The importance of selecting research stimuli: A comparative study of the properties, structure and validity of both standard and novel emotion elicitation techniques

Alexandra Caterina Constantinescu

PhD Psychology
The University of Edinburgh

2018

Contents

Abstract

The principal aim of this doctoral research has been to investigate whether various popular methods of emotion elicitation perform differently in terms of self-reported participant affect - and if so, whether any of them is better able to mimic real-life emotional situations. A secondary goal has been to understand how continuous affect can be classified into discrete categories - whether by using clustering algorithms, or resorting to human participants for creating the classifications. A variety of research directions subserved these main goals: firstly, developing data-driven strategies for selecting 'appropriate' stimuli, and matching them across various stimulus modalities (i.e., words, sounds, images, films and virtual environments / VEs); secondly, comparing the chosen modalities on various self-report measures (with VEs assessed both with and without a head-mounted display / HMD); thirdly, comparing how humans classify emotional information vs. a clustering algorithm; and finally, comparing all five lab-based stimulus modalities to emotional data collected via an experience sampling phone app. Findings / outputs discussed will include a matched database of stimuli geared towards lab use, how the choice of stimulus modality may affect research results, the links (or discrepancies) between human and machine classification of emotional information, as well as range restriction affecting lab stimuli relative to 'real-life' emotional phenomena.

Declaration

I hereby declare that this thesis is my own composition and that it contains no material previously submitted for the award of any other degree. The work reported in this thesis has been executed by myself, except where due acknowledgement has been made in-text.

Signed,

Alexandra Caterina Constantinescu

Acknowledgements

I would like to begin by thanking my family for their support and continual encouragement throughout this study period, and for making my aspirations and education their first priority even if this meant putting their own plans on hold. I would also like to thank my partner Mr. David Farrell, for his unwavering emotional support and confidence in my work, as well as the help he generously offered with some of the programming required by my research.

I would especially like to thank my supervisors, Dr. Sarah E. MacPherson and Dr. Adam Moore, for their advice and all the enjoyable meetings we spent discussing research plans and ideas. I highly appreciated being able to attend these supervisory meetings which I left filled with enthusiasm and excitement for the work I needed to do next. I am also indebted to Sarah and Adam for their feedback on my writing, and for challenging my ideas and forcing me to either discard or find more support for them.

Dr. James A. Russell's inspirational work made me interested in emotion research in the first place, so I feel I also owe him a lot of gratitude. I first came across his work during my undergraduate degree, and since then, the topic of emotion research became an interest of mine. I later had the pleasure of meeting him personally, and would also like to take this opportunity to thank him for his suggested reading, and the numerous interesting chats we had surrounding affective science. Since that same period of undergraduate study, another important influence on my current scientific thinking and the person who contributed most to my enjoyment of methodology and statistics, is Dr. Marian Popa. His lectures and advice (which I still rely on even now!) are something I treasure, and am most grateful to have had access to.

Dr. Donal Stewart also provided considerable programming assistance over the course of my doctoral research, and was also a model of uncompromising integrity when aspiring to high quality work. I am also grateful to Dr. Austin Tate for his suggestions on how to use Second Life for research.

Dr. Aja Louise Murray and Dr. Tom Booth made this journey so much more enjoyable for me, and in them I found two friends I value highly. Their company is something I take pride in, and I have been very fortunate to be able to share my passion for `R` with them, ever since they were organising the PsychStats meetings within the

Psychology Department.

I would also like to thank Brian Hills, Gillian Docherty, Dave Fitch, Dan Montgomery and everyone else at The Data Lab, for their understanding and friendly support during the final stages of writing up this thesis. Their patience and encouragement meant a lot to me, and I consider myself very fortunate to have them as my colleagues.

Last, but certainly not least, I would like to thank all my research participants, including my office mates who participated in my studies (in no particular order): Angela de Bruin, Thalia Theodoraki, Alicia Forsberg, Ilaria Taglialatela Scafati, and Anna Temp. Many others at the University of Edinburgh have also supported this body of doctoral research - unfortunately too many to name, which is something I must apologise for, and hope that they will understand.

List of Figures

List of Tables

Part I

Research context

Over the course of time, psychology has developed into a science with dedicated instruments, methods, and terminology. Among its numerous topics of enquiry, the study of emotions is not a recent endeavour - with some early attempts made in Antiquity by the Stoics (e.g., Posidonius or Chrysippus), or Aristotle and Plato (Sorabji, 2000)[1].

However, interest in the *scientific* study of emotions only started to peak in the late 19^{th} and 20^{th} centuries, which coincides, interestingly, with the abandonment of loaded terms such as "passions", "sentiments" or "affectations" in favour of the more neutral, and secular category of "emotions" (Dixon, 2003). Pivotal points along this path were, for instance, Darwin's work on the basic nature of emotions and the genetic origin of their associated facial expressions (work originally published in 1872), and James' perspective (1884) on how emotions follow physiological changes in the body in response to various stimuli, and represent our subjective experience of these changes (later known as the James-Lange theory, James, 1884; Lange, 1885).

Nonetheless, it would appear that the gradual transition to the term of "emotions" may not have been as big a step forward as one might hope, given the rich variety of phenomena which were now ascribed to this one term:

> *"Emotion is too broad a class of events to be a single scientific category. As psychologists use the term, it includes the euphoria of winning an Olympic gold medal, a brief startle at an unexpected noise, unrelenting profound grief, the fleeting pleasant sensations from a warm breeze, cardiovascular changes in response to viewing a film, the stalking and murder of an innocent victim, lifelong love of an offspring, feeling chipper for no known reason, and interest in a news bulletin. The boundaries to the domain of emotion are so blurry that it sometimes seems that everything is an emotion. The experts do not agree on what is an emotion and what is not. To be sure, all the different sorts of happenings included within this grab-bag term are important, some vitally so, but it is becoming increasingly clear that not all of them can be accounted for in the same way."* (Russell & Barrett, 1999, p. 805)

As a consequence, a unified and widely accepted theory of emotional phenomena is still lacking - perhaps with little hope of this being resolved in the near future, despite sustained efforts over the last two centuries, and constantly refined theories and research methods.

Given the context above, this thesis would like to set forward a provocative thought for consideration: how much of the ambiguity and difficulty surrounding the study of

[1] For an overview of the philosophical roots of the study of emotions, see Solomon (2008).

emotions might be attributed to the challenging nature of the topic itself, and how much perhaps due to inappropriate research methods that could be improved?

The rest of this chapter will explore this question by reviewing frequently used methods of eliciting, and then measuring emotions in psychological research. It will also consider alternative methods for eliciting emotions, and discuss these in the context of their "ecological validity" and their potential to (reduce) bias (in) results, relative to more typical methods.

Chapter 1

Emotion theories, elicitation and measurement

1.1 Theories of emotion

THE various methods of eliciting emotions and how they differ from one another will constitute a focal point in the current research. Generating emotions is a necessary process for research - given that in order to study an emotion, one must first witness it occurring. But due to the considerable number of procedures available for this, a legitimate, but extremely challenging, question is: What constitutes an "emotion", and when can we be certain that we have generated one?

Tentative answers vary by author and their adopted theoretical perspective. A rather harsh, but insightful critique of the current scientific / theoretical understanding of emotions is offered below:

> *"Many states generally regarded as important human emotions that form the stuff of plays, novels and garden-gate gossip are long term largely dispositional states, including jealousy, vengefulness, family love, obsessive ambition, infatuation, fascination with a mathematical problem, etc. There are other long term affective states such as preferences, and attitudes that are not normally called emotions. Of course, someone who defines an emotion as an episodic state in which there are particular sorts of bodily changes or sensed changes in a body state map will not include some of these long term states as emotions. But that's just another example of the terminological disarray."* (Sloman, 2004, p. 2)

With this criticism in mind, let us briefly discuss some widely-cited theories of emotion, using a classification based on work by Coppin and Sander (2010), and Prat (2006).

The account below is by no means complete, and could easily have formed the topic of (historical) doctoral research on its own, however it will serve to set the context for the current research.

1.1.1 Evolutionary theories

These theories have as their starting point Darwin's 1872 perspective, according to which emotions are universal, genetically pre-programmed modules with evolutionary significance, and which are also easily recognisable via prototypical facial expressions. As such, these facial expressions are said to have a signalling function for the underlying *discrete* emotional states, and an adaptive role within the environment.

Probably the best example of a theory following from this tradition, is the work carried out by Ekman and Friesen (1971) in New Guinea. Responses collected from an indigenous tribe (i.e., the Fore people, with little to no contact with the Western world) suggested facial expressions of emotions may be universal, given the relatively successful recognition rates of Western facial expressions by the indigenous population tested. This fuelled a strand of research into so-called "basic" (or discrete) emotions, considered to be: shared across cultures and possibly even species, relatively short-lived, spontaneous, and automatically triggered under specific conditions (e.g., sadness, as a basic emotion, is triggered by the death of a loved one). Furthermore, these basic emotions (i.e., usually anger, fear, happiness, sadness, disgust, and surprise - although the number can vary by author) also have adaptive value and therefore, evolutionary significance: disgust, for example, is considered to have evolved to protect individuals from harmful substances.

Under this framework, the existence of more complex emotions is accepted, however they are seen as combinations of these basic types. For instance, Plutchik (2009) who is also influenced by these ideas, divides emotions into primary and secondary, according to a circumplex model - with secondary emotions having less clear functions for survival and representing combinations of the primary ones: outrage, as an example, can be seen as a combination between the basic emotions of surprise and anger.

One example of how Darwin's legacy further impacted on modern thinking is the somatic marker hypothesis proposed by Damasio, Everitt, and Bishop (1996) (see also: Bechara & Damasio, 2005). Interestingly, here emotions are studied and understood via their relationship with decision-making, and are considered to have a pervasive influence on how individuals respond to stimuli. This is achieved by attributing various "somatic markers" (i.e., signals arising from bio-regulatory processes) to the various stimuli, which aids individuals in selecting an adaptive course of action for survival.

Attempts at defining emotions have been made not only based on the processes they influence, but also based on the temporal sequence in which they occur relative to other processes. A famous example is James' view (1884) on emotions as being the *outcome*, rather than the *source* of physiological changes arising in response to a stimulus (e.g., someone is not trembling because they are afraid, but rather, they feel fear after/because they are already trembling). This theory, however counter-intuitive, has some basis in reality, for instance in the case of fear, where a neural system exists to by-pass the neo-cortex, and directly link the thalamus and the amygdala. This enables a fast, adaptive response - while further processing might take longer, and only later produce the conscious emotion of "fear" (LeDoux, 1998). Under this perspective, emotions are seen as being of a "peripheral" nature, as well as related to specific patterns of physiological activation, which determine which emotion has been elicited.

Another theory influenced by this frame of thought is the facial feedback hypothesis (foreshadowed by work by Tomkins, 1980), which proposes that displaying full configurations of prototypical facial muscle movements (i.e., forming facial expressions associated with basic emotions) leads to a higher subjective feeling for the corresponding emotion. For instance, producing a facial expression engaging all the muscles typically associated with an honest smile, leads to heightened positive ratings when watching movies (Soussignan, 2002).

Cannon (1927) and Bard (1928), later contradict this "peripheral" view of emotions, by arguing instead for a "central" function of emotions, i.e., a stimulus will first be processed at the level of the central nervous system, and only then the individual's physiology will adjust accordingly. This perspective also claims that emotions cannot be clearly distinguished based on the pattern of physiological activation they induce. Somewhat confusingly, this is a valid point, with some empirical support (see, e.g., Christie & Friedman, 2004; Mauss, Levenson, McCarter, Wilhelm, & Gross, 2005). Another argument levelled against James' theory is that non-emotional states may also produce patterns of activation otherwise seen alongside affective processes (e.g., disgust may lead to physical feelings of nausea, but nausea can also appear spontaneously without emotion, following improper diet). Hence, the pattern of physiological activity is not an unequivocal indicator of emotion.

1.1.2 Cognitive theories

The roots of these theories can be found in Arnold's work (1960), where the key term of "appraisal" is introduced. This refers to a judgement passed by an individual to a stimulus or situation, which precedes and determines what type of emotional processing should ensue. This judgement can be based on criteria such as the novelty of the stimulus / situation, its relevance for the individual's goals, its compatibility with social or personal norms etc., and leads to nondescript affect or activation levels being labelled as discrete states (e.g., anger or fear etc.).

Another precursor for these theories is the work done by Schachter and Singer (1962), which draws some parallels with both the James-Lange and Cannon-Bard theories of emotion discussed previously. Namely, Schachter and Singer (1962) regard physiological activation as being necessary for the occurrence of an emotion, however they do not regard this physiological activation as having to present any specific pattern. Rather, in their view, the determining factor for which emotion is elicited, is the social context or environment the individual is in. This perspective will become known as the Two-Factor Theory of emotions.

Following from this tradition, further appraisal theories were proposed by Scherer and Ekman (2009), and the late Frijda (1986), the former of whom decomposed emotional episodes into five elements: the cognitive evaluation process, the level of physiological activation, a behavioural or motor component, motivational or goal-oriented aspects, and the subjective experience of the emotional episode. To this perspective, the latter author added more emphasis on the role of emotions as agents preparing individuals for actions (in the form of "action tendencies").

1.1.3 Social theories

These theories deny that evolution and genetics are defining influences on emotions, in favour of individuals' social environment (J. Averill, 1980; J. R. Averill, 1983; J. R. Averill & Nunley, 1988). Thus, emotions are born and interpreted based on the social context in which they arise, according to social norms, roles and expectations, and are only seen as forms of social responses.

For instance, according to this view, rather than being a basic emotion, anger fulfils a sophisticated social role in correcting wrongs and re-establishing moral order after a line has been crossed *intentionally* by a perpetrator (Cornelius, 2000). By additionally perpetuating the shared social understanding that an angry victim can behave *out of control* (i.e., without the calculated *intention* of harming the perpetrator), the societal balance can be restored, without the angry person having to bear the same kind of responsibility when retaliating, as the original perpetrator when committing the offence.

Similarly, fear can constitute a much more complex emotion than the evolutionary perspective would permit, when considering that e.g., fear of natural predators is but one narrow instance of fear, which also includes the fear of associating with people that society disapproves of, or fear of public embarrassment etc.

1.1.4 Psychological construction / dimensional theories

This perspective relies on the idea that emotional experiences can be described using a few continuous dimensions, and has its roots in early work by Wundt (1896) - who labelled these dimensions as Pleasure, tension and reassurance / inhibition. Subsequently, this view on emotions has been investigated further by Osgood (1952); Osgood, Suci, and Tannenbaum (1964) who re-labelled these dimensions as: evaluation, activity and potency, based on the results of factor analyses carried out on pairs of opposing verbal stimuli (e.g., fast-slow, hot-cold etc.). Mehrabian (1970); Mehrabian and Russell (1974) also extended the relevance of these dimensions for explaining non-verbal communication (e.g., posture, facial expressions etc.) and how common human scenarios are interpreted semantically and emotionally.

One notable influence on dimensional theories of emotion has been the work carried out by Russell and Mehrabian (1977), which further suggests that the dimensions:

- Pleasure-Displeasure (otherwise known as "Evaluation" / "Valence"),
- Degree of Arousal (or "Tension" / "Activity"), and
- Dominance-Submissiveness (or "Reassurance" / "Potency"),

are necessary and sufficient to adequately describe emotional states. Together, these three dimensions (Valence/Pleasure, Arousal, and Dominance) can also be referred to as the PAD model.

Just a few years later, Russell (1978; 1980) published additional results indicating that these dimensions are valid and do not depend on the research methodology used. His work also suggests that these dimensions are bipolar, with the Valence and Arousal acting as the axes of a two-dimensional space, and forming a circumplex around which emotional experiences are distributed. Various circumplex models of how these two axes interact in two-dimensional space have since been proposed (see Feldman Barrett & Russell, 1998; Yik, Russell, & Barrett, 1999).

Given this empirical work and the competing theories on emotions which were described previously, Russell (2003, 2005); Russell and Barrett (1999) developed a theoretical distinction between *core affect* and prototypical episodes of emotion. The former represents a consciously-accessible (though usually not salient), free-floating, and ever-present emotional tone, which involves varying levels of Valence and Arousal. It can morph into prototypical emotions, once a trigger-event has occurred. These display finite duration, and engage cognitive, physiological and behavioural reactions. The basic emotions which form the focus of Ekman's view are thus incorporated into this wider perspective, as prototypical emotional episodes (e.g., fear or anger). Similarly, appraisals are also integrated into this view, as a component within these prototypical emotions.

1.2 Criticism of emotion theories

As a comparison between the different views on emotion, a good starting point would be to consider the shortcomings of each, which are schematically outlined below. Based on this brief critique, one theoretical perspective on emotions will be adopted and used throughout this thesis:

1.2.1 Evolutionary theories / Basic emotions

- Inconsistencies in the number of basic emotions proclaimed over time, sometimes by even the same author (Cornelius, 2000): from four (Ekman, Levenson, & Friesen, 1983), six (Ekman, Friesen, & Ellsworth, 1982), or seven according to Ekman and Friesen (1986), eight according to Plutchik (1980), ten by Izard (1971), to as many as 22 according to Ortony, Clore, and Collins (1990); Ortony and Turner (1990). While the scientific process can legitimately lead to disagreement and constructive debate without this being a theoretical weakness per se, these discrepancies between the number of basic emotions may be too extreme to be explained in this manner;

- The method of studying how basic emotions are recognised very often relies on forced-choice items, which inflate the appearance of consistency (and thus, the amount of support for these theories), because participants implicitly assume the "correct answer" has to be among the options presented. Thus they are more likely to use a process of elimination to match the faces to the emotion labels provided (DiGirolamo & Russell, 2016; Nelson & Russell, 2016);

- School children become increasingly better at recognising facial expressions of emotion over time - suggesting this is a learned response, rather than an innate capability. At very young ages, they are only able to distinguish coarse, Valence-based differences in expressions, i.e., positive vs. negative expressions, without improvements in specificity even when using dynamic over static faces as stimuli (Widen, 2013; Widen & Russell, 2015). Hence, if basic emotions are indeed evolutionarily old, these theories cannot explain this differential progression in the ability to recognise them;

- The universality hypothesis cannot explain the cultural differences in how diagnostic information is sampled from the face in order to distinguish surprise from fear, or disgust from anger, i.e., Eastern cultures oversample information from the eye area, which are very similar between surprise and fear on the one hand, and disgust and anger on the other. Hence, this culture is more likely to confuse these pairs of "basic emotions" (Jack, Blais, Scheepers, Schyns, & Caldara, 2009; Jack, Garrod, Yu, Caldara, & Schyns, 2012). In addition, research on indigenous populations with little to no contact with Western societies, has been repeated and shows modest recognition levels for these "basic emotions" (Crivelli, Jarillo, Russell, & Fernández-Dols, 2016; Crivelli, Russell, Jarillo, & Fernández-Dols, 2017);

- Over-reliance on facial expressions as unequivocal indicators for the associated emotional states leads to limitations and confusing scenarios for emotions which may not have any particular facial expression (e.g., guilt and shame). Similarly, different emotions may share the same facial expression (e.g., joy and love may both share the smile, Posner, Russell, & Peterson, 2005);

- Reliable connections have yet to be established between basic emotions and patterns of physiological activation - which would contradict the notion that basic emotions represent different, and easily distinguishable modules (Cacioppo et al., 2000; Christie & Friedman, 2004).

1.2.2 Cognitive theories

According to Coppin and Sander (2010), Moors (2009) and Moors, Ellsworth, Scherer, and Frijda (2013):

- It has been hotly debated whether cognition (e.g., appraisal, among other processes) is necessary for emotions to occur. However, appraisal theorists have eschewed this question by conceding that *conscious* cognition may not be needed for emotional processing, but that *unconscious* cognition is instead, as a bare minimum. This assertion entails empirical difficulties - not the least of which is operationalising the term "cognition", when it is such an all-encompassing term. For instance, deciding what meal to have, estimating the distance between two locations, evaluating the potential consequences of a bad grade, inhibiting the urge to eat sweets when on a diet, or remembering to phone an old friend, could all count as forms of cognition. This makes the claim of (unconscious) "cognition" being necessary for emotions to occur, extremely difficult to falsify.

- In addition to the above, empirical findings such as the mere exposure effect (Zajonc, 1968, 2001) suggest that appraising a stimulus in a certain way, even nonconsciously, is not necessary for emotional processing - given that merely perceiving the same stimuli repeatedly outside the focus of consciousness leads to evaluating them in a more positive light (i.e., attributing higher Valence to them).

- There is inconsistency across authors in terms of the number and identity of appraisal variables needed for an emotion to arise, i.e., criteria on the basis of which the appraisal occurs (e.g., novelty, goal-relevance and so on). For instance in the case of anger, it is debatable whether a necessary component of the appraisal should be *agency* / blame (Moors, 2009). Similarly, some cognitive theories include Valence as an appraisal variable, but not others (Moors et al., 2013). A further issue which is debated is whether these appraisal variables represent categorical or continuous dimensions, which remains an open question.

- There is disagreement about the temporal sequence of emotional sub-processes, and whether they should occur sequentially or in parallel. However, one more popular proposition would propose the sequence to be:

  `Stimulus` → `Appraisal of stimulus` → `Action tendency` → `Physiological responses` → `Behaviour` → `Attribution / Labelling of emotion`

  (Moors, 2009; Scherer, 2001)

  Regardless, in this case appraisal theories do not respect empirical findings such as LeDoux (1998); Schachter and Singer (1962), given that the cognitive component is considered to occur just after perceiving the stimulus, and prior to physiological responses - instead of allowing the possibility for the physiological response to precede the cognitive treatment.

- Using the excitation transfer paradigm (Cantor, Zillmann, & Bryant, 1975; Tannenbaum & Zillmann, 1975; Zillmann, Katcher, & Milavsky, 1972), it is possible to show that arousing activities can enhance the emotional response for unrelated stimuli, if their onset partially overlaps with the original activity (i.e., any residual activation from the initial activity gets transferred onto the second). Thus, participants can mis-attribute their total level of activation to the wrong cause - which also poses problems for cognitive theories positing that each stimulus is appraised in some way before any emotion can arise (e.g., Scherer, 2001; Scherer & Ekman, 2009).

- In the case of phobias, even when individuals are explicitly aware that the object of fear is actually harmless, the sense of irrational fear remains (Griffiths, 2004). This would also contradict the importance of appraisals as part of emotional processing. Interestingly, it may also be the case that phobias are influenced by genetic factors, which appraisal theories are also unable to account for (Kendler, Myers, Prescott, & Neale, 2001);

- Despite the belief that appraisals need not be conscious, a popular, but contradictory, method for investigating appraisals is participant self-report.

1.2.3 Social theories

According to Prat (2006), these theories rely on explaining emotions based on transient social roles and socially prescribed ways to react. Thus, the influence of internal processing on the part of the individual is largely ignored, although social norms are understood and responded to by means of stimulus evaluations, and interpreting cognitive schemas, which the individual *is* responsible for.

Aside from the obvious vacuum in explaining the role of the individual's own psychology in modulating his/her own emotional states, these theories can also be criticised for failing to explain how or why an individual can fake an emotional response solely because it is desirable within the social context, but hide their real emotional state - given that it is the *real* emotional state which these theories aim to explain (Ekman, 2009).

1.2.4 Psychological construction / dimensional theories

- In a two-dimensional model (Valence × Arousal), qualitatively different emotions could occupy the same location in the circumplex, e.g., fear and anger (because both are negative and can be high in Arousal). However, this criticism can be fairly easily overcome when adding the third continuous dimension of Dominance (in our example, anger would prototypically be associated with high Dominance, whereas fear, with low Dominance).

- There is some ambiguity in terms of how Arousal is interpreted: as intensity or activation. The former would represent just a property of Valence, whereas the latter would be an separate dimension, varying from low activation (i.e., relaxation, boredom, sleepiness), to high activation (i.e., nervousness, being jittery, or excited etc.).

- These theories do not leave any margin for considering the potential adaptive value which emotions may have, or at least this aspect of emotions is not considered by these theories.

- Core affect may be more of a "theoretical" construction, which cannot easily be studied empirically.

Clearly, based on all the shortcomings associated with each theoretical view and described above, the study of emotions is a complex process, and is still a work in progress. And yet, we find that the dimensional view presents particularly attractive attributes:

- According to Eerola and Vuoskoski (2010), by combining several continuous dimensions, this view can more adequately characterise emotionally ambiguous stimuli, relative to the basic / discrete emotions view which focuses on just a subset of "stereotypical" cases of emotions, e.g., anger or fear.

- It appears to capture more closely the essence of how emotions function, relative to the competing models. For instance, Lindquist, Gendron, Barrett, and Dickerson (2014) have shown that with progressing cases of semantic dementia (SD), basic emotions / discrete emotion categories become less recognisable, in contrast to variations in core affect as assessed using continuous dimensions. SD patients prove to be less able to use conceptual knowledge of emotions (e.g., "anger" or "sadness"), but are still able to distinguish whether an emotion is positive or negative in Valence. This suggests that the dimensional approach more closely represents the deeper structures underpinning emotions and how they operate, since such an approach is still able to accurately predict/explain experimental results even when the influences of verbal labelling and conceptual knowledge are removed.

- Posner et al. (2005) also present convincing evidence for the validity of this model, in terms of self-report, physiological / biological, behavioural and brain data. For instance, when exploring relationships between these different channels, behavioural measures such as zygomatic and corrugator muscle activity were found to align with self-reported measures of Valence; skin-conductance measures were also shown to be associated with self-reported Arousal, etc.

- The theoretical framework for the dimensional view on emotions can be conveniently paired with a measurement model and tools, e.g., the Self-Assessment Manikin (Bradley & Lang, 1994).

- This view (relative to, e.g., the social construction view), supports easier quantification of emotional stimuli (e.g., the IAPS image database has been normed using this framework, Lang, Bradley, & Cuthbert, 2008, 1999) and comparisons between studies (Fontaine, 2009).

For all these reasons, the dimensional view of emotions will be adopted in the current work - both in terms of the measurement paradigm used for empirical research, as well as for the interpretative framework for discussing results.

1.3 Eliciting emotions

The empirical study of emotions often relies on special methods or tasks designed to induce particular states. Subsequently, these states can be investigated and quantified - a process which emanates from, and feeds back into, emotion theory. What is often less clear is how much impact the choice of elicitation method may have on empirical results, and therefore, on how theories of emotion advance over time.

1.3.1 Overview

Attempts have been made at classifying these methods, according to whether or not they involve participant deception, whether participants are aware what emotion is intended to be generated, or whether the induction procedure involves the participant completing a task alone, or by interacting with other persons (Kappas & Descôteaux, 2004). However, this is a challenging task, given the wide variety of emotion elicitation / induction techniques. Nevertheless, a few common choices[1] are, according to Coan and Allen (2007); Gilet (2008); Martin (1990); Stemmler (2003):

Words / Text. Two known examples of word / text databases are: The Affective Norms for English Words (ANEW; Bradley & Lang, 1999a), and The Affective Norms for English Text (ANET, Bradley & Lang, 2007a). Such stimuli have easily been used especially to study the interaction between emotions and other cognitive processes. However, given that these are particularly prone to cultural / language effects, they often require validated versions depending on the country where research will take place, e.g., Stadthagen-Gonzalez, Imbault, Pérez Sánchez, and Brysbaert (2017) for Spanish words, and Võ et al. (2009) for German.

Velten technique. This method relies on the participant reading (silently, and then out loud) a sequence of 60 self-referential statements, intended to induce either positive or negative states, e.g., "This is great - I really do feel good. I am elated about things.", vs. "Things are easier and better for other people than for me. I feel like there's no use in trying again.". This method was first proposed by Velten (1968), and others have developed this work since, e.g., Richardson and Taylor (1982), but this method has remained a topic for debate, due to the possibility of high demand characteristics.

Sounds. Separate strands of research have opted for this type of stimuli, depending on

[1] Despite reflecting different processes, some emotion *and* mood induction procedures will be considered here, given that it is difficult to clearly distinguish between them (i.e., When using these techniques, can we definitely say that we have elicited a mood, and not an emotion, or vice versa?). In addition, some of these techniques can often be used for both purposes, e.g., music.

whether the point of focus was the affective processing of environmental sounds (Thierry & Roberts, 2007), or human voices (Belin, Zatorre, Lafaille, Ahad, & Pike, 2000). Similarly to film clips discussed below, a multitude of options exist for selecting audio stimuli, including affective databases such as the International Affective Digitized Sounds (IADS) - versions 1 (Bradley & Lang, 1999b) and 2 (Bradley & Lang, 2007b), or freely available online collections of audio files.

Music. According to Gilet (2008); Martin (1990), this method was first used by Sutherland, Newman, and Rachman in 1982, and relies on either participants or researchers using selected pieces of music to reach a certain emotional state. It is also often associated with other types of emotional stimuli (e.g., imagery or text, Mayer, Allen, & Beauregard, 1995). A variation of this method uses pieces of music that are identical at the beginning, but later develop into either joyful, sad or neutral tunes, in the attempt to lower demand characteristics (Martin, 1990).

Images. Static visual stimuli are very widely used in research. Perhaps the most popular example is The International Affective Picture System (IAPS; Lang et al., 2008), i.e., a collection of several hundred affective images, varying in Valence, Arousal and Dominance, as well as content type. A similar, but more recent, example is The Nencki Affective Picture System (NAPS, Marchewka, Żurawski, Jednoróg, & Grabowska, 2014). Other image databases may focus specifically on facial expressions - see R. Gross (2005) for a listing.

Film clips. These are some of the most commonly used types of stimuli for inducing emotions, particularly in controlled environments. In addition, according to Westermann, Spies, Stahl, and Hesse (1996), films are also one of the most effective such methods. Despite this, Bartolini (2011) claim that only three film clip validation studies were conducted prior to 2011, surprisingly. This may be related to the fact that it is relatively common to create one's own selection of film stimuli, rather than opt for extant stimulus databases, which are more likely to be associated with larger-scale validation work. Nonetheless, several film databases exist, such as those constructed by J. J. Gross and Levenson (1995), A. Schaefer, Nils, Sanchez, and Philippot (2010) and Carvalho, Leite, Galdo-Álvarez, and Gonçalves (2012).

Autobiographic recall. This method was proposed by Brewer, Doughtie, and Lubin in 1980. Typically, participants are asked to close their eyes and recollect several past events (sometimes in writing), as clearly and in as much detail as possible. These may be either negative or positive events that have happened to them. According to Gilet (2008)'s review, this method is particularly effective for eliciting positive states. On the other hand, given that participants are a key-component of the procedure, the objectives of which are easy to discern, this method may also be prone to demand characteristics irrelevant to the emotional aspect(s) of the research.

Guided imagery / imagination. This method relies on presenting participants with various scenarios / short stories, and asking them to imagine that they are in the situations described (e.g., Bond, 1998; Miller, Patrick, & Levenston, 2002). Such fictional situations have the advantage of introducing more standardisation than, e.g., autobiographic recall, but may also suffer from demand characteristics, similarly to previous techniques.

The Directed Facial Action Task. This method is closely related to the Facial Feedback Hypothesis mentioned previously, and consists of producing facial muscle contractions that resemble expressions of the universal emotions. This is thought to trigger the associated feelings in participants, even when they are not instructed to pose a certain emotion, but rather asked in a neutral fashion to contract various facial muscles (Levenson, Ekman, & Friesen, 1990).

Success-failure manipulations. This method consists of providing a task for participants to perform, and regardless of their real rate of success, they are falsely informed that they have failed or succeeded on the task, depending on whether the intention is to induce a negative or positive emotional state, respectively. Examples are bogus intelligence tests, sport performances, or exam feedback. A meta-analysis for this elicitation method was conducted by (Nummenmaa & Niemi, 2004), and discovered that this method appears valid and useful regardless of whether it aims to induce positive or negative states. However, they do caution that as a pre-requisite, participants must be invested in the task given, in order for success or failure to be regarded as important.

Social psychological methods. This method typically relies on the use of confederates and temporary deception of participants. For example, anger can be elicited using interpersonal insult, in a design where participants are asked to write about their personal values and beliefs, and then tricked into thinking that a second participant (which actually does not exist), has read their material and based on it, thought very poorly of them and their level of education. A strength of this method is that it is perhaps more suitable to elicit such emotions as anger, compared to e.g., images or films (Harmon-Jones, Amodio, & Zinner, 2007; Harmon-Jones & Sigelman, 2001; Lobbestael, Arntz, & Wiers, 2008).

Virtual reality. Even though used less frequently than the methods described previously, Virtual Reality (VR) has also been used to elicit or modulate emotional states (see e.g., Baños et al., 2006), and shows promise due to some unique features (see the next section, and Section 9.2.4, p. 387 for details). Other popular elicitation methods have various methodological disadvantages: e.g. demand characteristics when reading sentences aimed to induce a given emotion; difficulty in standardising real-life manipulations with confederates; eliciting different emotions to the varying degrees (Lench, Flores, & Bench, 2011); or producing inconsistent results across different age categories (e.g. the same film clip may efficiently elicit disgust for a young sample but not an older one, Kunzmann, Kupperbusch, & Levenson, 2005).

As a response to these concerns, VR could offer more realistic stimuli and a higher degree of experimental control and replicability (G. Young, 2010) - which otherwise seldom (if ever) occur all three of them simultaneously. We believe that these positive features have led to the increasing popularity of VR, which we have attempted to capture in Figure 1.1, part (a), on page 48.

1.3.2 Comparing elicitation methods

Which of the elicitation methods described above is the "best" will depend on researchers' aims (Kappas & Descôteaux, 2004), however, in order to make an informed choice, extensive research is needed to investigate if there are any systematic differences between all these methods, and precisely which of them is most suitable for which research design and question.

Yet, despite the implications of this topic for research methodology and the interpretability of results, it appears that relatively little research has been done in this area. Even though a few examples exist of comparative studies, they usually represent smaller-scale studies, contrasting only a few elicitation methods simultaneously. For instance, Salas, Radovic, and Turnbull (2012) compared internally-generated (i.e., via autobiographic recall) and externally-generated (i.e., via film clips) discrete emotions, and found that they are largely equivalent in terms of their ability to induce the intended emotion, and their self-reported levels of intensity - with the exception of joy, where the internal procedure was superior.

Uhrig et al. (2016) also researched this area and compared two external techniques: images and film clips, whereas Jallais and Gilet (2010) compared autobiographical recall with a composite technique involving music and guided imagery (i.e., essentially imagining various scenarios such as: "It's your birthday and your friends throw you a terrific surprise party", alongside a congruent musical background). Both studies yielded somewhat surprising findings, particularly for the Uhrig et al. (2016) study, where images proved both more arousing and more useful for triggering the intended states than films.

As for the Jallais and Gilet (2010) study, autobiographic recall was found to be superior to the composite technique for inducing departures from participants' baseline emotional states.

The study led by Zhang, Hui, and Barrett (2014) was more ambitious in terms of the number of elicitation methods compared - four: two composite techniques both of which included music, combined with one of either autobiographic recall, or images. Guided imagery and posing affective facial expressions, postures and vocalisations were also added to these. The two composite techniques (i.e., music, with either images, or autobiographic recall) outperformed the other two methods in terms of generating positive states (which shares some degree of similarity with the findings of Salas et al., 2012 mentioned above), but only the composite method involving music and images was also effective at generating negative states.

The meta-analysis conducted by Lench et al. (2011) includes a section on comparing various types of emotion elicitation methods (i.e., films, images, priming, music, Velten, imagery, text, behavioural techniques including the Directed Facial Action Task, social psychological methods, and autobiographical recall). A key finding is that overall, images constituted the method associated with the largest effect sizes, surpassing even film clips, despite these being used more frequently in research, according to the authors. The other methods were associated with moderate effect-sizes, with the exception of priming, which was the least effective method.

When assessing if any / which elicitation types may be more suitable for specific emotional states, the results are less clear: films and images were consistently associated with large effect sizes when distinguishing induced happiness from other non-target emotions: sadness, anger or anxiety, but both failed to register more than a moderate effect size when contrasted to a neutral state. Films maintained this pattern of small to moderate effect sizes for all other contrasts (e.g., sadness vs. anger, anger vs. anxiety etc.), and so did images, with two exceptions where they again led to relatively large effect sizes: states of anger or anxiety, both contrasted with neutral states. Other elicitation methods showed more inconsistent patterns, which seemed to differ based on the type of emotion induced, and what it was compared against. For instance, behavioural methods (e.g., the Directed Facial Action Task) were particularly useful for eliciting anger, when compared to either a neutral or happy state, or anxiety (when compared to a happy state), but otherwise generated moderate effects, at best, for all other contrasts.

Given the inconsistent findings in this area, exploring these effects is worth further research, as well as assessing the potential of VR for inducing emotions. In fact, in Figure 1.1 (page 48, constructed using data gathered from Google Scholar, with Publish or Perish software, Harzing, 2017), we show that a steady increase has been occurring in terms of the publications related to both Virtual Reality and emotions, when in parallel, other methods are decreasing in popularity.

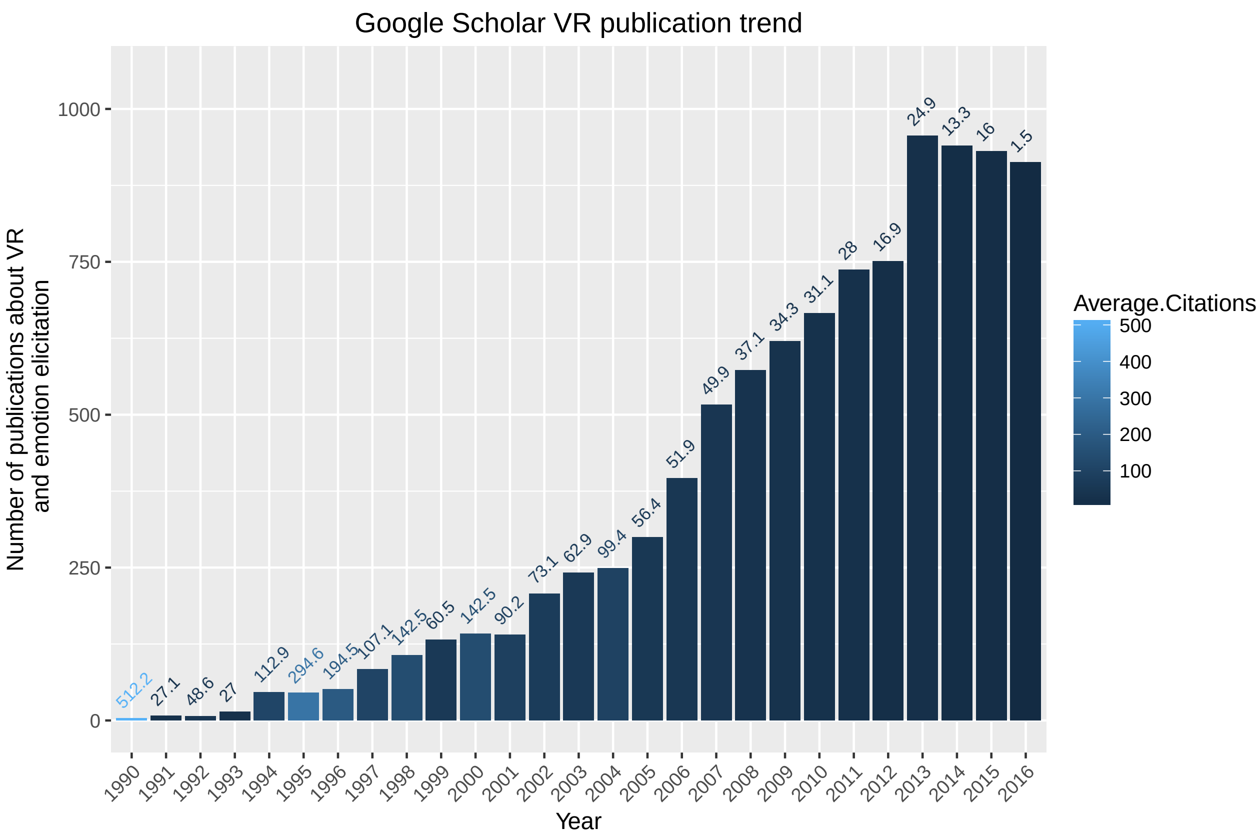
The same pattern was observed by Fox, Arena, and Bailenson (2009), who also provide a fairly comprehensive account on the usefulness of VR for the social sciences. Even ten years previously, Loomis, Blascovich, and Beall (1999) had presented a similar perspective on the uses of VR as a research tool in psychology. Both articles discuss how VR can create more realistic testing conditions, which still retain adequate experimental control. In fact, manipulations which are relatively ineffective or even impossible with other means become a reality in VR, e.g., inducing fear by placing participants' in front of a virtual precipice instead of having them read a story about this, or altering participants' own race to investigate social stereotypes. Additionally, VR has also proven to be an effective means to elicit emotions, so much so that it can counteract states produced by other forms of emotion elicitation (such as Velten, Bouchard, 2010).

McCall, Hildebrandt, Hartmann, Baczkowski, and Singer (2016) make a similar point regarding "ethological" / "cognitive ethology" approaches to research (e.g., Kingstone, Smilek, & Eastwood, 2008), where processes should be studied within the complex environments in which they are meant to operate, instead of the impoverished and artificial conditions from research labs. In a study where participants are allowed to freely wander around several virtual environments, the authors found that exploration patterns and gaze directions in VR largely reflect behaviours expected in real-life, e.g., angry-faced avatars in virtual rooms are avoided by participants, but positive objects in these rooms are instead approached. In a room with sudden and scary events, participants tended to show freezing behaviours and less visual exploration of the environment - again, a result that seems natural and in agreement with expectations for behaviour in real-life surroundings.

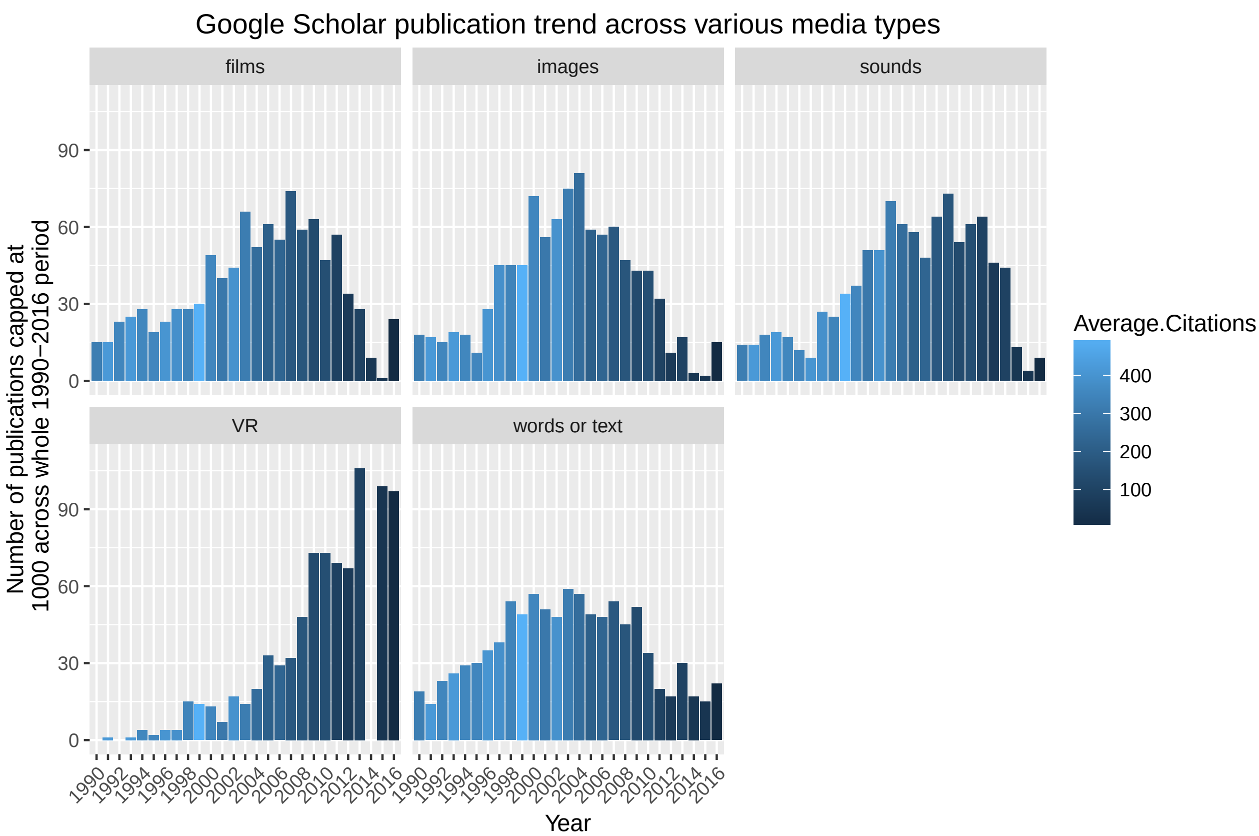
Kuliga, Thrash, Dalton, and Hölscher (2015) investigated how participants respond to a real building vs. a VR representation of the same building, in terms of various self-report dimensions[2]. Few differences were found between the two, and these mostly concerned "atmospheric" differences (i.e., lighting conditions in VR, relative to real-life). The reason for this is assumed to be the failure to have incorporated an advanced lighting model into the virtual environments, including some form of ambient occlusion (i.e., casting shadows in a scene in a realistic way that takes into account the geometry of the entire scene and multiple light sources). This is a good indication that with the advancement of computer graphics, it will become possible to blur the distinction between real and virtual, enough to make use of the latter in research very reliable.

[2] Arousing/Calming; Bare/Decorated; Boring/Interesting; Coherent/Incoherent; Cold/Warm; Illegible/Legible; Inaccessible/Accessible; Light/Dark; Monotone/Varied; Narrow/Spacious; Novel/Familiar; Open/Closed; Private/Public; Scary/Relaxing; Simple/Complex; Ugly/Beautiful; Unattractive/Attractive; Unclear/Clear; Uninviting/Inviting; Unpleasant/Pleasant.

Taken together, these findings point to VR as a promising avenue for research into emotions, which requires further validation against other more typical methods of elicitation, as well as real-life situations.



(a) ``virtual reality'' AND ``emotion'' AND (elicit* OR induc*). Search conducted on Dec. 22, 2016



(b) one of the following: (words OR text) / images / sounds / films / ``virtual reality'', followed by: (emotion OR affect) (elicitation OR induction) (method OR technique). The the phrase "in other words" from the word / text was excluded from the search, conducted on Dec. 22, 2016.

Figure 1.1: Rise articles published between 1990-2016 in Google Scholar, for the search queries under each plot. Queries capped by Google at 1000 results per search. Data compiled using **Publish or perish** software (Harzing, 2017).

1.4 Measuring emotions

The measurement of emotions can be carried out at four levels of inquiry: self-report, physiological, behavioural and neural. Triangulating measurement methods in this way can in theory provide valuable insights, but in practice, frequently these four streams do not provide a coherent picture of affective processes, particularly for healthy samples (see Mauss et al., 2005; Mauss & Robinson, 2009; H. S. Schaefer, Larson, Davidson, & Coan, 2014).

Nielsen and Kaszniak (2007) suggest that a possible reason for this disagreement may be the inadequate methods for measuring self-report, which fall short of capturing the true richness of personal experiences. Even assuming they could, they would be unable to do so in a very concise way, so as not to force participants to reflect on their experience too much during measurement, and hence risk modifying it. This may induce low correlations between self-report measures and the other three streams of information[3].

Nonetheless, the same authors also suggest that emotional self-report can be a more stable source of inquiry than is often thought. For instance, results reported by Burton and Kaszniak (2006) suggest that while facial muscle activity for emotional responding is impaired in patients with Alzheimer's disease (i.e., a behavioural component of emotions), their self-reported emotional state still remains similar to that of healthy controls after viewing emotional images.

In addition, self-report may also be more reliable when participants' attention is directed towards dimensions already known to organise emotional experiences (i.e., Valence and Arousal), compared to completely open-ended descriptions of emotional states, where participants may fail to direct their attention to such components of their experiences. Self-report methods also tend to be the only easily accessible and affordable method for capturing states in naturalistic settings (e.g., using experience sampling techniques outside research labs). For these reasons, and despite some reservations, the validity of self-report measures need not be an "all-or-nothing phenomenon" (Mauss & Robinson, 2009).

1.4.1 Self-reported measures of emotion in the current work

In order to gain a better understanding of how self-report can be used to measure emotions, Zentner and Eerola (2010) have provided a useful classification of these methods, which we have reproduced in Table 1.1. Of these, the methods ***highlighted*** are those

[3] However, the AffectButton (to be discussed shortly) is a novel self-report method which can assess multiple emotional dimensions very quickly, thus increasing the likelihood of measuring "raw phenomenality", rather than reflective post-modifications of the experience

which we have employed in future chapters of this thesis. We also employed a similar method to the non-verbal evaluation task, where emotional stimuli were classified according to their similarity, and given verbal labels freely by participants.

Table 1.1: Table adapted from Zentner and Eerola (2010) listing various measures of self-report.

	Instrument	Description
1	***Likert Scales***	Likert-style ratings of emotion concepts
2	Adjective checklist	Selection of appropriate adjectives
3	Visual analogue scales	Continuous rating scales without intermediate steps
4	***Continuous response versions of self-report instruments***	Continuous evaluations of emotion concepts using a computer interface
5	Non-verbal evaluation tasks	Arrangement of emotional stimuli according to their similarity, without use of verbal labels
6	***Experience sampling method***	Structured report of ongoing activities related to emotion and their causes at times prompted by a pager
7	Diary study	Detailed daily report of the central emotional episodes and theirs causes and effects
8	Free / phenomenological report / narrative method	Description of the personal experience. The actual format and focus may vary greatly (retrospective reports over a lifetime of experiences, writing about the recent important emotional episodes, etc.)

In order to investigate emotional processing, we believe we have used a representative sample of these methods, i.e., (non-verbal) Likert scales (the Self-Assessment Manikin), continuous-response devices (the AffectButton), experience sampling, and a free stimulus classification / sorting task, based on the perceived affective similarity between our stimuli. We will further describe some of these methods below.

1.4.1.1 The Self-Assessment Manikin, or SAM

Bradley and Lang (1994) have applied the dimensional view on emotions to the construction of a now widely-used measurement tool: the Self-Assessment Manikin (or SAM). Using five expressive cartoon figures, each scale (i.e., Valence, Arousal and Dominance) is represented non-verbally, making this instrument particularly attractive for use across different cultures. Participants are instructed to either select a cartoon figure or the space in between two neighbouring figures, to indicate their emotional state with reference to a particular stimulus etc.

Thus, by using both the figures and the spaces between them, three 9-point scales are proposed. Others have further developed the tool, by adding more figures to each scale, and thus removing the need to mark the space between adjacent figures as a response, e.g., Irtel (2008); Suk (2006). An example is presented in Figure 1.2.

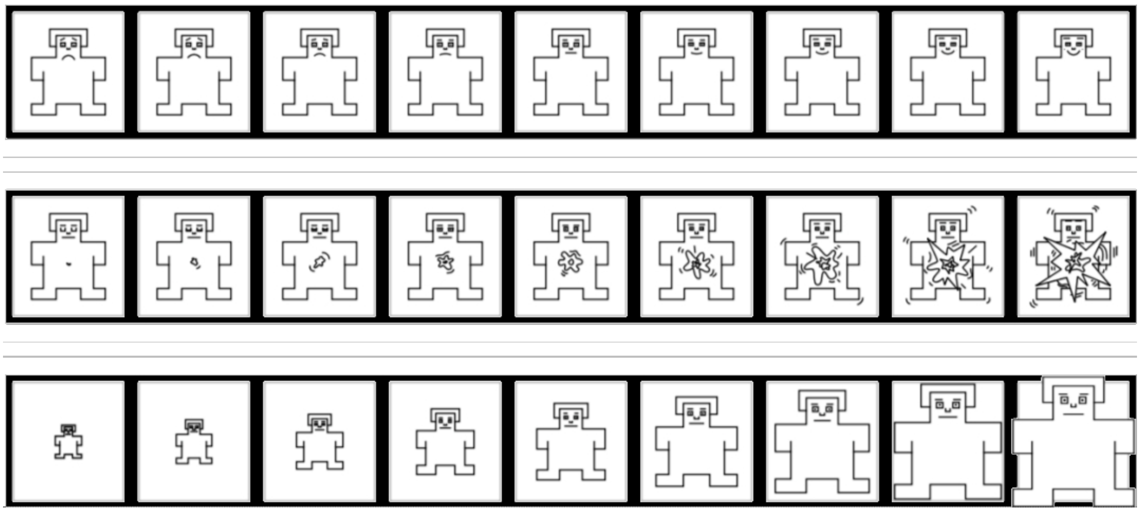


Figure 1.2: The Irtel (2008) SAM scales.

1.4.1.2 The AffectButton

In terms of self-reported emotional states, according to Ruef and Levenson (2007), variations in affect can be measured continuously over time, using a tool called the Affect Rating Dial. Simply put, this instrument typically presents itself as a rotating switch which can be turned right or left to indicate how positive or negative a given emotional experience is, and how this changes over time.

However, one obvious criticism would be that this tool measures only one dimension of affect: Valence. Even if the ability to measure this continuously over time is extremely useful, there remains no possibility to discern between various types of negative or positive affect (e.g. on the one hand: grief or panic etc. and on the other hand: relief or excitement etc.).

The Affect Grid (Russell, Weiss, & Mendelsohn, 1989), however, serves precisely this purpose, by allowing participants to simultaneously provide a Valence and Arousal rating. The Affect Grid is reproduced in Figure 1.3, and relies on participants writing an "×" in a cell of their choice.

However, the question arises about how to build an affect rating dial able to measure all three dimensions of affect at the same time according to the PAD model - also without overwhelming participants. A method and code to achieve this have been proposed by Broekens and Brinkman (2009): the Java applet created by the authors uses a flattening of a 3D- onto a 2D-space, to measure concurrent variations in Valence, Arousal and Dominance (Mehrabian, 1995).

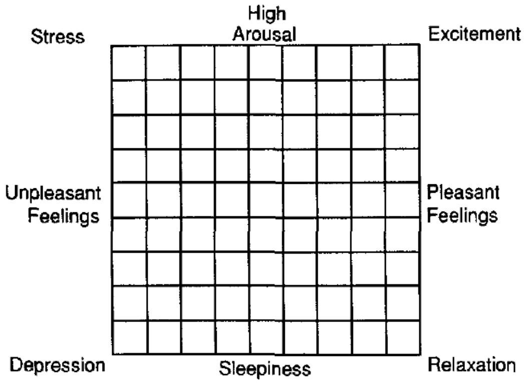


Figure 1.3: The Affect Grid (Russell et al., 1989). Vertical variations represent changes in the level of Arousal, whereas horizontal changes reflect variation in Valence. Diagonals describe simultaneous changes on both dimensions, e.g., the top right corner corresponds to high Pleasure and high activation, i.e., excitement.

This is achieved in a very intuitive manner, by using a caricature placed within a square, which changes its facial expressions according to where participants' mouse[4] hovers over it - see Figure 1.4 on page 53. Movement along its x and y axes controls Valence and Dominance, and moving closer to the square edges defines the level of Arousal. When participants identify a facial expression which they think suitably represents their current emotional state, they can submit this as their rating during an experimental task. The data saved by the applet with each submission is the numeric set of Valence, Arousal and Dominance coordinates controlling the facial expression selected; a screenshot of the graphical expression itself is not saved.

1.4.2 The influence of measurement conditions, and representative experimental design

Areas of research where experiments typically represent *abstract versions* of how processes really occur in real life, risk obtaining results that no longer address the initial research question, but rather just the abstract laboratory conditions themselves (see Dhami, Hertwig, & Hoffrage, 2004; Jerit, Barabas, & Clifford, 2013; Kingstone et al., 2008; Mitchell, 2012). Emotion research seems to be an area particularly prone to this type of situation, considering studies may approach issues such as emotional memory across the life-span, by asking different age groups to recall emotional word lists of varying Valence (Kensinger, 2008). It is unclear how well such an artificial situation would generalise to real-life, where participants seldom, if ever, are required to memorise disconnected word lists.

[4] or finger, for touchscreen devices.

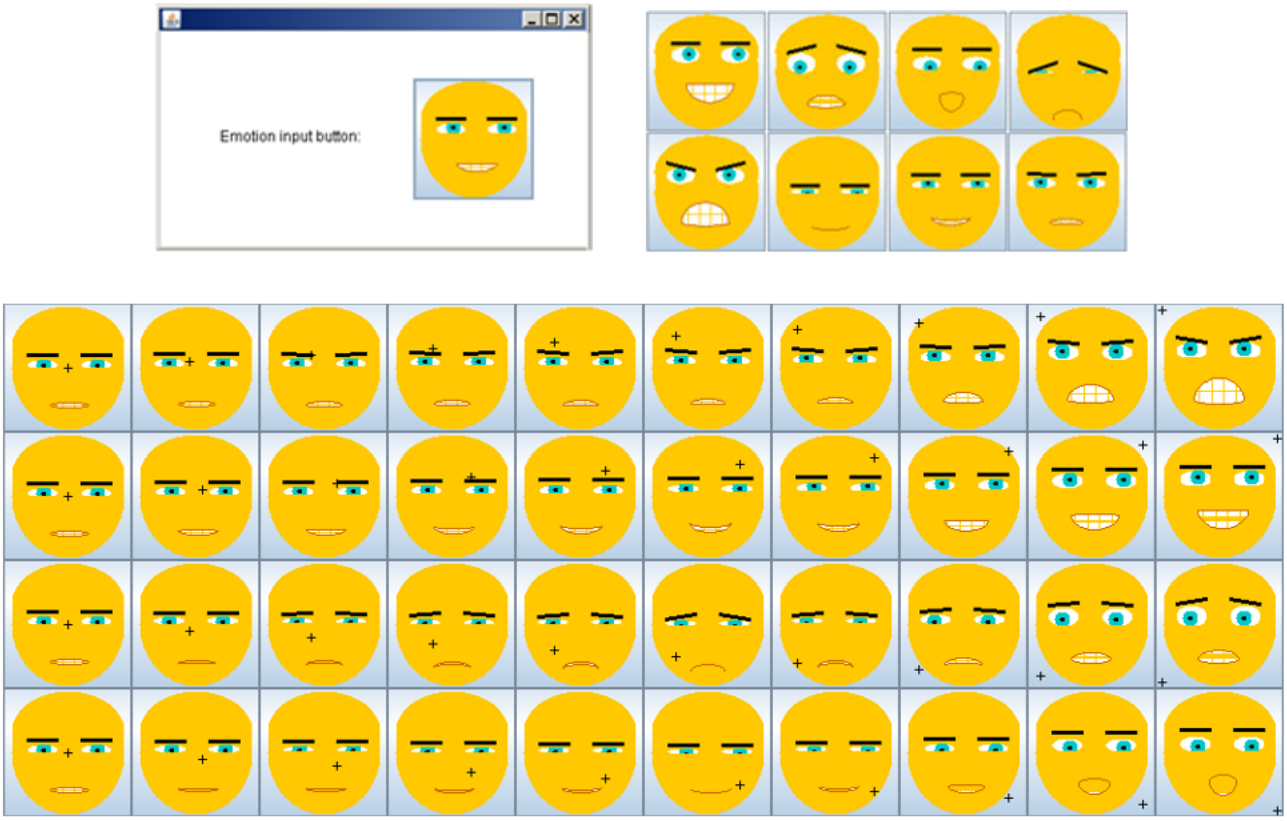


Figure 1.4: The AffectButton (Broekens & Brinkman, 2013). The variety of expressions represents the translation of the cursor position (displayed as a + sign) into Valence, Arousal and Dominance scores which control the face.

Another such example is the study by Mather and Nesmith (2008), where the topic of inquiry was the extent to which feature binding is influenced by emotion. To investigate this, participants were required to undergo 144 trials, each of which consisted of: watching a single image briefly displayed in a given location on a computer screen, which would then vanish, and then indicating whether a dot that had newly appeared on the screen instead was yellow or green. After having finished all the trials, participants again watched the images presented previously, except this time, three copies of any given image would appear simultaneously on the screen, only one of which would be in the same location seen previously. Participants were asked to indicate which of the three copies occupied the same position as before. They then reviewed all the pictures again, and rated them in terms of Valence and Arousal. It is immediately apparent that such a task never occurs in daily life, and that this might perhaps explain the contradictory findings which exist in this area, namely that stimulus Arousal either impairs or, to the contrary, supports successful feature binding, i.e., of stimulus identity and location (Mather, 2007).

Last but not least, studies of facial expression recognition may also fall prey to the same issue, given that the recognition of emotion in daily life presumably uses both facial cues, voice, posture and gesture changes. However, in research labs, facial expression recognition tasks often rely only on static images, which are of faces only. Such studies have concluded, e.g., that there is an age-related decline in the capacity for recognising emotions from facial expressions (Isaacowitz et al., 2007). However, the simple addition of auditory cues to static facial expressions - making them more similar to how they would occur in daily-life - seems to reduce age differences (Hunter, Phillips, & MacPherson, 2010). This too underscores the importance of using more naturalistic testing conditions in research.

As further support, Jerit et al. (2013) are some of the few authors to have compared laboratory to field settings in research, with results indeed suggesting that findings differ across settings. Causal mechanisms for this are thought to relate to the temporal closeness between the treatment and measurement of the dependent variable in the lab (which can artificially inflate effect sizes that are otherwise modest in the real world). Additionally, in some cases it is obvious what is being measured in the lab, which might also boost effect sizes. Such elements are usually absent in the "outside world", making it unlikely to witness the same relationships whenever behaviours are observed naturally.

To our knowledge, the first to consider at length the importance of how measurement conditions affect generalisation was Egon Brunswik, whose work developed the theme of sampling not only representative *individuals* from the population of interest, but also representative *experimental procedures* for how psychological processes actually operate in their natural environment (Hammond, 1948).

As such, Brunswik (1955) claimed that psychology should be a "science of the organism and its adaptation to the environment", and not merely a science of the organism, taken on its own. If it were to achieve this goal, he suggested that the widely used *systematic* designs (i.e., the typical, natural-science inferential approach, where only the treatment is allowed to vary, while all other interfering variables must be held constant) should be abandoned in favour of more comprehensive designs.

Even if the popularity of systematic designs (which include factorial designs) has made them almost synonymous with the experimental method *per se*, they constitute an attempt to separate variables orthogonally (which is both unrealistic and conducive to artificial stimuli), instead of allowing the natural covariation between variables to be maintained (for a more detailed explanation, see the excellent work by Dhami et al., 2004). This natural covariation is referred to as "the causal texture of the environment", Tolman & Brunswik, 1935), which is likely to be very revealing in itself. This is because, according to this perspective, psychological processes are seen as adaptations to the environment in a Darwinian sense (Hammond, 1966). Hence, attempting to study them in a void via systematic designs, empties them of meaning.

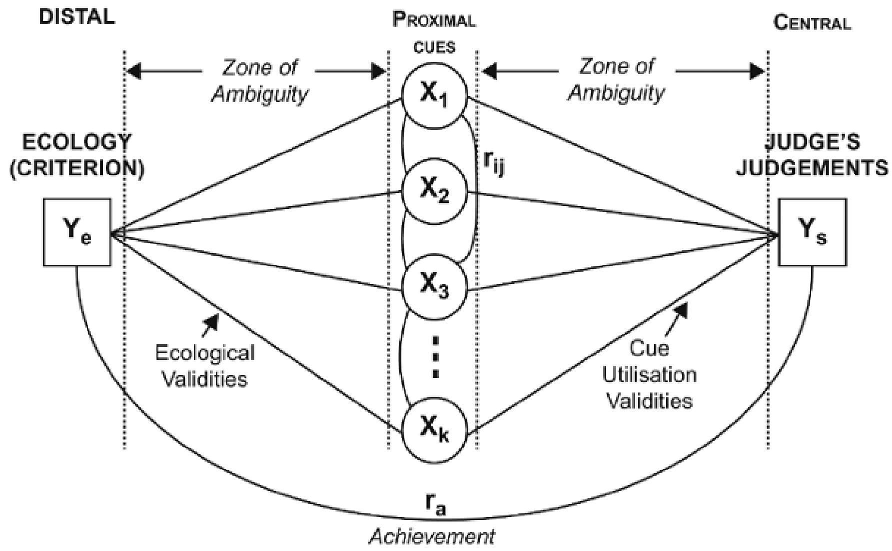


Figure 1.5: Brunswik's lens model[5]**.** In their process of adapting to the environment, individuals often only have access to various proximal indicators, which they need to use as "clues" to infer the state of a distal variable. These clues may or may not be used in a valid manner - but if they are indeed used successfully, they will approximate the true "ecological validities" between those cues and the distal criterion.

These ideas have been incorporated into Brunswik's "lens model", reproduced in Figure 1.5. The main function of this model is to illustrate the *probabilistic laws* that govern an individual's adaptation to his environment, i.e., the successful use of available (proximal) cues in order to infer the state of an unavailable (distal) object (Dhami et al., 2004). In order to achieve this, the individual must estimate the *ecological validities* of the proximal cues, or the correlation each such cue bears with the distal variable, that the organism means to achieve. As such, it becomes obvious that in its original use, the term "ecological validity" does **not** refer to the similarity between a controlled setting and a natural one (Araujo, Davids, & Passos, 2007; Dhami et al., 2004; Dunlosky, Bottiroli, & Hartwig, 2009; Hogarth, 2005). Instead, Brunswik (1955) had devised an altogether different term to designate this: **representative experimental design**.

While being very suitable for describing heuristic-based decision-making, this model might also provide insight into the functioning of emotions. For instance, if someone is lying to us (i.e., the distal variable), we might try to tell by using various cues / proximal variables, e.g., if the person makes eye-contact, is breathing too fast or contradicts him-/herself. It is possible to use such cues correctly and successfully identify the lie (because the "ecological validities" of those cues were correctly estimated), or we might use the cues sub-optimally, if we did not also consider that the person is perhaps shy, which might also account for his/her behaviour. Depending on which interpretation was selected (the

[5] Image source: unknown.

other person is a liar vs. simply shy), we might feel anger or maintain a relatively neutral state. This model permits to explicitly incorporate the notion of *probabilistic rules* into emotional processing - which current theories frequently omit. In doing so, this may explain some of the inconsistencies in the field and the large variation in emotional phenomena.

Additionally, this model can (potentially) allow a better understanding of emotions, also by allowing them to be studied using the associated research methodology, instead of typical methods. Hence, as opposed to the vague and widely (mis-)used term "ecological validity", the idea of representative design is associated with various methods of measurement and designs. According to Dhami et al. (2004) and Hammond (1966)[6], two major options are available:

A. Substantive situational sampling: Refers to the actual content of the task and is congruent with Brunswik's original definition of representative design. Several sub-types exist:

1. **Random / probability sampling:** This includes time sampling or the experience sampling methodology (Csikszentmihalyi & Larson, 2014; Scollon, Kim-Prieto, & Diener, 2003), where each stimulus has an equal probability of being selected, under the assumption that with enough cases sampled for measurement, an interpretable central tendency will emerge.
2. **Stimulus canvassing / Non probability sampling:** Relative to the previous method, non probability sampling does not afford cases the same chance of being sampled, hence this is considered to be a rather primitive method. Subtypes here include stratified, convenience, or typical-case sampling. These are attractive for practical reasons, but do not allow wide generalisations.
3. **Complete coverage of stimulus population:** A method previously deemed unrealistic, where data is being measured continuously in the natural environment. However, with the advances in technical equipment (e.g., sensors, mobile phone geolocation tracking etc.) and methods for dealing with big data, it appears more and more reasonable to aim for this.

B. Formal situational sampling: If Brunswik himself was advocating sampling real stimuli directly from the environment, Hammond also saw the possibility of stimuli that could be constructed with the pre-assessed, formal properties of the environment, in mind. Hence, this method focuses on devising situations that reproduce the formal properties of the natural environment (e.g., number and values of cues, their intercorrelations, etc.).

[6] Who was Brunswik's student.

Dhami et al. (2004) specifically mention Virtual Reality as a possibility for implementing formal situational sampling, however with the caveat that prior to this, the natural ecology must be studied first. This provides a further argument that VR is a much-needed improvement in research, as it could implement this type of design.

1.5 Research aims and questions

Based on the emotion theories, elicitation and measurement methods previously discussed, the current work will attempt to investigate one major topic: whether the choice of elicitation technique can alter participant self-reports, and therefore have implications for distorting emotion theory.

This overall topic will be deconstructed into several strands of research:

a. Do affective words, sounds, images, film clips and virtual environments differ in their ability to elicit emotions? If so, in what ways? Due to design considerations (i.e., potential use for formal situational sampling, and higher affordances for reflecting the "real" world), Virtual Reality (VR) was regarded as a point of interest, in terms of its capacity to induce emotions in lab-settings. In order to further determine its potential relative to other already widely-used methods, four points of reference were chosen: affective words, sounds, images and film clips.

Other than their popularity, these four additional stimuli types were also selected due to formal similarities with VR: they all represent "external" elicitation methods, unlike, e.g., autobiographic recall, which would have been less suitable a term for comparison. Additionally, these methods can be seen as following a progression regarding the amount or type of information they convey to participants, and/or the degrees of freedom they afford to them.

Thus, words can act as a control for non-verbal stimuli such as images and sounds, and films can be seen as a natural progression from images and sounds, since they convey both an audio and visual stream simultaneously. Finally, VR may itself be viewed as one step further than films, in terms of additionally allowing participants to interact with situations, rather than just observe passively.

This line of inquiry will be addressed between Chapter 2 and Chapter 6, as follows:

Chapter 2: This chapter focuses on using the popular International Affective Picture System (IAPS), and and providing an objective way to systematically selecting stimulus groups from it based on the norms available for this database.

Chapter 3: Here we provide a data-driven method for sampling affective words (ANEW database) and sounds (IADS-2 database) to match the already chosen IAPS images. The outcome will be a matched list of stimulus trios, each containing a word, sound, and a corresponding image.

Chapter 4: After a brief introductory passage (see beginning of **Part III**), in this chapter the stimulus trios will be compared against a set of films, and only the most appropriate films will be retained as matches for the previous stimuli. Hence, the output of this procedure will be stimulus quartets, matched as closely as possible and including: one word, sound, image and film clip.

Chapter 5: A sub-selection of the film clips from the previous chapter will be tested against a group of virtual environments (VEs), in terms of the self-reported dimensions of Valence, Arousal and Dominance. This measurement model is the common link across all the stimulus types, from the international stimulus databases for words, sounds and images, to the films and virtual environments specially selected for this research.

Chapter 6: In this study, self-report data was collected using a subset of the VEs, and a Head-Mounted-Display (HMD). This was carried out to test whether the level of immersion used with the VEs affected previous results.

b. How do humans classify emotional information, relative to statistical algorithms?

Chapter 7: Relative to previous chapters, this line of inquiry will use an additional facet of self-report: stimulus classifications. This will serve to investigate how participants process continuous emotional information to form discrete categories, and whether this process can be mimicked by cluster analysis algorithms using the Valence, Arousal and Dominance dimensions. Discrepancies between human behaviour and "rational", statistically optimal algorithms will be interesting for creating insight into how humans (do not) perform this task.

c. How do these elicitation methods compare to daily-life experiences?

Chapter 8: This chapter will address whether data collected using a substantive sampling procedure via a phone app, can serve to validate any of the previously-discussed elicitation methods, particularly VR.

Finally, **Chapter 9** (the General discussion) will summarise the major findings identified across all the previous chapters, and situate them within the wider literature, organised by themes.

Part II

Selecting affective stimuli

As discussed in the previous chapter, numerous forms of emotion elicitation have been developed for use in research on emotions, from various types of affective media (e.g., images, films etc.), to tasks with confederates, imagination or autobiographical recall, to name but a few.

Some of the most widely used stimulus databases originate from the same research lab, and include:

- The International Affective Picture System (IAPS; Lang et al., 2008);
- The Affective Norms for English Words (ANEW; Bradley & Lang, 1999a);
- The International Affective Digitized Sounds (second version, by Bradley & Lang, 2007b);
- The Affective Norms for English Text (ANET, Bradley & Lang, 2007a).

The following chapters will focus on establishing a standard method to select stimuli from each of these databases, beginning with the IAPS in Chapter 2. This same chapter has also been published separately within a peer-reviewed journal (see Constantinescu, Wolters, Moore, & MacPherson, 2016). The publication is included within Appendix H.1 (for the accepted manuscript), and Appendix Section H.2 (for the supplementary material).

This work on the IAPS database will then be extended in Chapter 3, where we propose a strategy for sampling matched stimuli from all the other databases mentioned. The ultimate aim of this body of work is to provide a set of stimulus quartets (i.e., groups composed of one word, sound, image and short sentence each), which can then be used in empirical research, as well as to subsequently inform the selection of stimuli from two new modalities: film clips and virtual environments.

Chapter 2

Study 1: Selecting affective images

2.1 Introduction

It is now widely accepted that emotion plays a critical role in human psychology and is inextricably entwined with behaviour and cognition. Yet, a major challenge that emotion researchers face is conceptualising the relationship between various kinds of emotions and mapping their collective impact on other psychological processes (e.g., Ito, Cacioppo, & Lang, 1998; Lane et al., 1997; LeDoux, 1998. Perhaps the most widely used tool in this pursuit is the International Affective Picture System (IAPS; Lang et al., 2008), which consists of 1,182 images and is designed for the experimental study of affective processing. It is based on the PAD model, involving Pleasure/Valence, Arousal, and Dominance - a three-dimensional framework for measuring emotions (Mehrabian, 1996; Russell & Mehrabian, 1977). The validity of this theoretical model has accumulated a wealth of empirical evidence over time, and the number of citations for the database and instruction manual is now approaching 3,300, indicating a continued and robust research community surrounding it.

Using the IAPS database is particularly attractive due to the large variety of stimuli offered, as well as the chance to replicate and compare findings more easily between studies. Following the PAD model, each complete IAPS case is associated with normative (average) ratings for Pleasure/Valence (i.e., how positive or negative an image is), Arousal (i.e., how alerting or calming an image is), and Dominance (referring to the viewer's perceived amount of control in relation to the stimulus displayed). To exploit the flexibility offered by such a large number of stimuli, several typical approaches for image selection have been used, with some of the most common being discussed below. However, it is important to note that most of these methods rely on assumptions about the underlying multidimensional structure of the database, and that violations of those assumptions can have profound consequences with respect to what inferences may be drawn from experiments using these stimuli. Specifically, if those assumptions are unsustainable, then some of the conclusions from the emotion literature may be questionable.

2.1.1 Typical methods for sampling IAPS stimuli

2.1.1.1 Establishing group cutoff points

This method consists of selecting cutoff values, which usually divide one of the three continuous PAD distributions into different categories. For instance, Mikels, Fredrickson, et al. (2005) distinguished between positive and negative stimuli on the basis of which IAPS images had Valence ratings above or below 5, respectively, given the rating scale used to measure PAD dimensions in the IAPS contains nine points. Similarly, Xing and Isaacowitz (2006) considered the images with Valence scores between 1 and 4 to be negative, those between 4 and 6 to be neutral, and those over 6 to be positive, with images very close to these cutoff points being excluded (Xing and Isaacowitz, personal communication, June 6, 2015).

A variant of using group cutoff points is selecting extreme groups of images. This consists of retaining the first n most negative/positive images (or an upper and lower group of images), as well as a group with minimal distances from what is considered a "neutral" score. For instance, one of the four types of emotion induction used in Zhang et al. (2014)'s study consisted of a combination of images and music, with some of the images being selected from the IAPS stimuli according to their rank (most positive, most negative, or most neutral).

Another extension of the cutoff point method was used by Lithari et al. (2010), who combined it with graphical presentation and selected images on the basis of how they were organised within a 2-D space. Four quadrants were formed through the crossing of the Valence and Arousal nine-point axes at a score of 5, and each quadrant was considered to represent a separate group of stimuli.

The cutoff point approach is best suited to research questions that focus on only one dimension of the PAD model. Although carefully chosen combinations of cutoff points may be adequate when a study focuses on only one or two dimensions, this strategy becomes unwieldy when researchers intend to systematically vary all three dimensions at the same time. Moreover, the use of cutoff points in this fashion tacitly assumes that the non-controlled dimension(s) has (have) no effect on information processing or behaviour that is relevant to the researchers' interests - an assumption that is risky at the best of times. Finally, another implicit assumption, for which there appears to be no clear evidence, is that the groups formed using the cutoff points can approximate the internal structure of the IAPS data correctly.

2.1.1.2 Discretisation and crossing / controlling dimensions

This method refers to cutting the continuous PAD dimensions associated with the IAPS into n categories. Subsequently, within one such category, one may repeat the procedure on the basis of the remaining dimensions. For example, after cutting Valence ratings into three categories, one may then attempt to find images of varying levels/categories of Arousal within, for example, the most pleasant Valence category. Alternatively, one may attempt to control one dimension within another - for example, finding one category with relatively constant Arousal within the most pleasant Valence category.

For instance, Tomaszczyk, Fernandes, and MacLeod (2008) chose IAPS stimuli on the basis of their Valence ratings, but in addition attempted to cross different levels of Arousal within the Valence categories (see also E. Anderson, Siegel, & Barrett, 2011). Similarly, de Arcos, Verdejo-García, Peralta-Ramírez, Sánchez-Barrera, and Pérez-García (2005) selected five categories of images for eliciting emotional experiences, including one neutral Valence category with low Arousal, and positive and negative Valence categories, each with either a lower or a higher Arousal level. Finally, Perri et al. (2014) divided the IAPS stimuli into positive, negative, and neutral categories based on their Valence scores, with the first two of these categories presenting high levels of Arousal. The neutral-Valence pictures were selected to present low Arousal.

If attempting to cross PAD dimensions in a factorial design in this manner, the assumption is made that the PAD dimensions are orthogonal (i.e., uncorrelated), which is not what the IAPS data suggest (Bradley & Lang, 2007c). Similarly, attempting to control dimensions assumes that groups of images exist within the IAPS that vary in terms of one dimension, but not another. This is also generally not feasible, given that the correlated PAD dimensions tend to vary *together*. Finally, as is the case when using cutoff points, this method cannot easily accommodate the use of all three PAD dimensions simultaneously, usually resulting in Dominance scores being ignored. Although it is correlated with the other two PAD dimensions, Dominance represents a distinct entity with- in the model, and thus can itself account for some variation in affective ratings (Bradley & Lang, 1994). Therefore, if Dominance scores are ignored, this variation would be excluded from the image selection process, which poses risks for its validity.

2.1.1.3 Content selection

This type of stimulus selection based on content is usually combined with one of the previously discussed methods. For instance, Bernat, Patrick, Benning, and Tellegen (2006) selected erotic and adventure scenes as pleasant, and violent or threatening images as unpleasant stimuli. Neutral images were chosen to portray common objects or inactive people, and so on. In addition, this strategy was combined with dimension discretization/crossing, leading to groupings of pleasant and unpleasant images with low, medium, or high Arousal levels (see also Tomaszczyk et al., 2008). In another study, Hamann and Mao (2002) selected IAPS images on the basis of their content: pleasant pictures were chosen to depict erotic scenes, food, or agreeable animals and children. Negative images were selected thematically to include mutilated bodies, violence, and so forth. In parallel, high-interest images included exotic parades and surrealistic scenes, and low-interest images included plants or household scenes.

In addition, Eizenman et al. (2003) emphasised the thematic selection of IAPS images: four categories were selected to include images considered neutral, dysphoric, threatening, or socially themed. However, the authors also relied on Valence ratings to guide their selection procedure, so that neutral images were selected to have Valence scores close to 5, threatening/dysphoric images ranged between Valence scores of 2 and 4, and the social themes presented a range between 6 and 8 on the same scale. They also aimed to control variations in Arousal levels by allowing maximum differences of two points across the images in each of the four categories. The content selection method does not place strong assumptions on the data on its own; however, it is usually used conjointly with the content selection, discretisation and crossing/ controlling dimension methods, which do.

2.1.2 An alternative image selection method based on cluster analysis

The present work offers an alternative strategy for image selection based on clustering algorithms, which can be used with all three PAD dimensions simultaneously. To our knowledge, such algorithms have been used to categorise participant responses from individual studies (e.g., for classifying brain regions with differential response patterns to disgusting vs. neutral images - Deen, Pitskel, & Pelphrey, 2011; or for grouping participants in terms of their risk for alcohol abuse, on the basis of heart rate variability in response to IAPS emotional stimuli - Mun, von Eye, Bates, & Vaschillo, 2008), but not to group or select images on the basis of normative data.

In this article, we argue that clustering methods constitute a valuable means for creating experimental stimulus groups based on the IAPS normative data, by ensuring that group formation is optimised according to various measures (e.g., maximising the distances between the different groups or the likelihood that cases belong to a certain group). This can boost the level of statistical power achieved in studies, since the larger the differences between levels of the treatment, the higher the chances of finding significantly meaningful effects (see Hallahan & Rosenthal, 1996, p. 495).

In addition to using more objective criteria for group formation, relative to entirely "manual" methods, clustering algorithms can also capture the particular structure of the IAPS data, and thus provide image classifications that are more empirically principled. This can allow experimenters to guard against confounds in the form of heterogeneous, systematically underpopulated, or "artificial" categories of stimuli, which cannot be adequately supported by the IAPS database. For instance, IAPS images are often divided into three groups based on Valence. However, if this three-group structure is not an adequate fit for the IAPS normative data, images may be grouped inappropriately. Thus, if multiple types of negative material exist within the IAPS, creating only one category of negative images would risk blending these together, with unpredictable consequences for study results and the validity of any inferences based on them.

In addition, without consulting the structure of the IAPS data (which clustering methods are sensitive to), it might be tempting to resort to a factorial design combining three ordered levels of Valence (low, neutral, and high) with as many levels of Arousal. In this situation, it would be difficult to find enough images populating the intersection between low Valence (i.e., negative images) and low Arousal (i.e., relaxing images), due to the correlation between these two dimensions. Indeed, such a category could thus be deemed "artificial", as it would ignore the essential correlations between PAD dimensions.

Consequently, clustering methods can provide information on both the quantity and quality of stimulus categories that can realistically be supported by the structure of the IAPS normative data. Although such algorithms can be flexibly adapted to extract a predetermined number of groups, usually they are allowed to follow an exploratory strategy constrained by the overall structure of the data set. That is, they will find the "best" number of stimulus clusters/groups, subject to some optimisation constraints. This is a point of departure from the typical selection methods discussed above, in which a top-down process is often used to identify three image categories fitting the notions of "negative", "neutral", or "positive". Finally, clustering algorithms can limit the amount of labour associated with stimulus selection, especially when research hypotheses involve more than one feature being taken into account at the same time (i.e., Dominance, as well as Valence and Arousal). Indeed, by minimising this difficulty, the method we propose below allows researchers to expand the scope and complexity of their hypotheses, and thus more easily test their theories.

Our hypothesis is that the IAPS data present a discernible, meaningful structure that can be capitalised upon by using cluster analysis to produce stimulus groups for experimental use. Here we tested several clustering approaches against one another, and propose a stepwise strategy for filtering and classifying IAPS images for subsequent experimental use. The family of clustering algorithms (or data-mining techniques) is extremely diverse and easily warrants entire books dedicated to them (for more detailed discussion, see Jain & Dubes, 1988; Kantardzic, 2011; Kaufman & Rousseeuw, 2005). However, due to their widespread use and popularity, we focus on several approaches in particular. We will now briefly describe each of these approaches; readers interested in a more in-depth coverage may refer to the supplementary material included with Appendix H.2 (from p. 562).

The first approach is ***k*-means clustering**, which involves selecting k random seeds (i.e., random points in the space defined by the dimensions of the stimuli) and assigning the closest cases to them, leading to the formation of k groups. Afterwards, the group mean (i.e., centroid) is computed, and cases are reassigned to groups on the basis of closeness to this value. This process will reiterate until the classification has settled into a stable solution (i.e., when the data points no longer change their memberships after the centroid computation). This is a hard partitioning method, meaning that all cases are included in their respective clusters with a probability of 1, and it does not provide a direct indication of the number of clusters existing in the data (Hartigan & Wong, 1979; MacQueen, 1967; Xu & Wunsch, 2009). Instead, various subsequent indices are used to suggest the number of clusters that would be appropriate for a given dataset. However, these do not take parsimony into account, and so may show little consistency or be prone to inflating the number of clusters. In order to establish clusters of images that could later be used as the levels of an "emotional content" independent variable, we tested k-means clustering because of its efficiency, simplicity, and wide use (Jain, 2010).

Another popular option is **hierarchical clustering**. This is an agglomerative method whereby individual cases begin by being designated as their own cluster (i.e., clusters of one data point each; Borcard, Gillet, & Legendre, 2011; Xu & Wunsch, 2009). Using one of multiple *linkage methods*, cases get merged progressively into ever-larger clusters, until all of the cases belong to just one, overarching cluster. Similarly to k-means, no indication is given about the suitable number of clusters in the data, so that with the aid of various statistical criteria (again not considering parsimony, and possibly conflicting in their recommendations), it is largely up to the re- searcher to decide where along this progression to stop and retain the corresponding number of clusters. Hierarchical clustering is also a hard clustering method, in which each case is assigned to one cluster exclusively, rather than being assigned a probability of membership.

A third option that is gaining in popularity is **model-based clustering**. This represents a form of hierarchical clustering that also involves an expectation-maximisation (EM) procedure (for a primer on EM, see Do & Batzoglou, 2008). Unlike k-means, or hierarchical clustering per se, this is a soft clustering method, whereby cases are assigned to clusters with a certain probability (uncertainty) of membership. This can allow researchers to systematically control for the degree of typicality a stimulus exhibits in terms of the clustering dimensions used: A stimulus with higher uncertainty will be less representative of its cluster, and may introduce additional noise into experimental results. Also, in contrast with the two previous approaches, model-based clustering simultaneously provides both a clustering solution for the data and a straightforward method for determining the optimal number of clusters k. For this purpose, model-based clustering (implemented in the `mclust R` package: Fraley & Raftery, 2006) provides Bayesian Information Criterion (BIC) values and considers the optimal number of clusters for a given dataset to be whichever value of k maximises[1] this criterion. Therefore, one of the distinctive features of this method is that it takes parsimony into account in the attempt to reduce the unnecessary inclusion of components (clusters) into the model.

To summarise, in this article we focus on three types of clustering - namely k-means, hierarchical, and model-based clustering - each of which differs in the approach taken to assigning case membership (and whether that membership is probabilistic or absolute). Moreover, the first two approaches do not intrinsically provide a clear criterion for determining the final number of clusters, and so admit a variety of methods for deciding this (see below, and in the supplementary material included within Appendix H.2). We tested each of these methods on the IAPS data in order to: (a) gain more insight into the internal structure of the database; (b) identify any common patterns in clustering solutions across the different algorithms; (c) select the most suitable algorithm of the three and retain its clustering solution, and lastly; (d) extract a fixed number of representative IAPS images from the final clustering solution for use in further experiments.

Subsequently, we employed various validation techniques, to select one clustering method as the most appropriate for the IAPS dataset. After selecting one such clustering algorithm, we extracted the best exemplars from each resulting cluster, which we then propose as the final selection of stimuli that researchers may wish to use in subsequent work.

2.2 Method

2.2.1 Dataset characteristics

In this study, we focused on the IAPS normative data gathered from both male and female participants, in which PAD ratings were collected using three (nonverbal) 9-point Likert scales (using the Self-Assessment Manikin, or SAM; Bradley & Lang, 1994; Lang, 1980) and a sample of approximately 100 US students, depending upon the image. In our analysis, we included all three PAD dimensions that are available within the IAPS data, to create stimulus groups that account for the maximum amount of variance in participant responses (Bradley & Lang, 1994). Despite the large correlations between

[1] The formula employed by Fraley and Raftery (2006) uses the negative of deviance, so that BIC here needs to be maximised rather than minimised, which is more common.

Dominance and the other two PAD dimensions (see Figure 2.1), Dominance did not perfectly overlap with them (e.g., if r ¿ .9) either empirically or theoretically, further justifying its inclusion in subsequent analyses.

2.2.2 Duplicates

We evaluated the univariate distributions available within the stimulus database, and identified 12 duplicate cases within the normative data (overall including N = 1,194 cases, but with N = 1,182 unique cases), each associated with different scores on the PAD model (see Table 2.1 for a listing) for a listing). These images were likely normed twice, as part of different image sets (Lang et al., 2008). As a consequence, we replaced these duplicated pairs with a single entry containing the averaged Valence Arousal, and Dominance across the duplicates.

Table 2.1: Table of IAPS duplicates and their Valence, Arousal and Dominance ratings (devised using the `Stargazer R` package by Hlavac, 2013).

Description	Image code	Valence	Arousal	Dominance
Spider	1230	4.090	4.850	4.580
Spider	1230	4.610	4.030	5.600
Horse	1590	7.180	4.740	5.540
Horse	1590	7.240	4.800	5.620
Rabbit	1610	7.820	3.080	6.770
Rabbit	1610	7.690	3.980	6.520
Coyote	1640	6.270	5.130	5.220
Coyote	1640	6.160	5.180	4.910
Cow	1670	6.810	3.050	6.530
Cow	1670	5.820	3.330	5.630
NeutFace	2210	4.380	3.560	5.030
NeutFace	2210	4.700	3.080	5.230
Mutilation	3000	1.450	7.260	2.990
Mutilation	3000	1.590	7.340	2.730
Mutilation	3010	1.710	7.160	2.880
Mutilation	3010	1.790	7.260	2.880
EroticFemale	4220	8.020	7.170	5.330
EroticFemale	4220	6.600	5.180	5.900
EroticMale	4520	7.040	5.480	5.480
EroticMale	4520	6.160	4.800	5.730
AimedGun	6200	2.710	6.210	3.350
AimedGun	6200	3.200	5.820	3.490
Exhaust	9090	3.560	3.970	4.510
Exhaust	9090	3.690	4.800	4.720

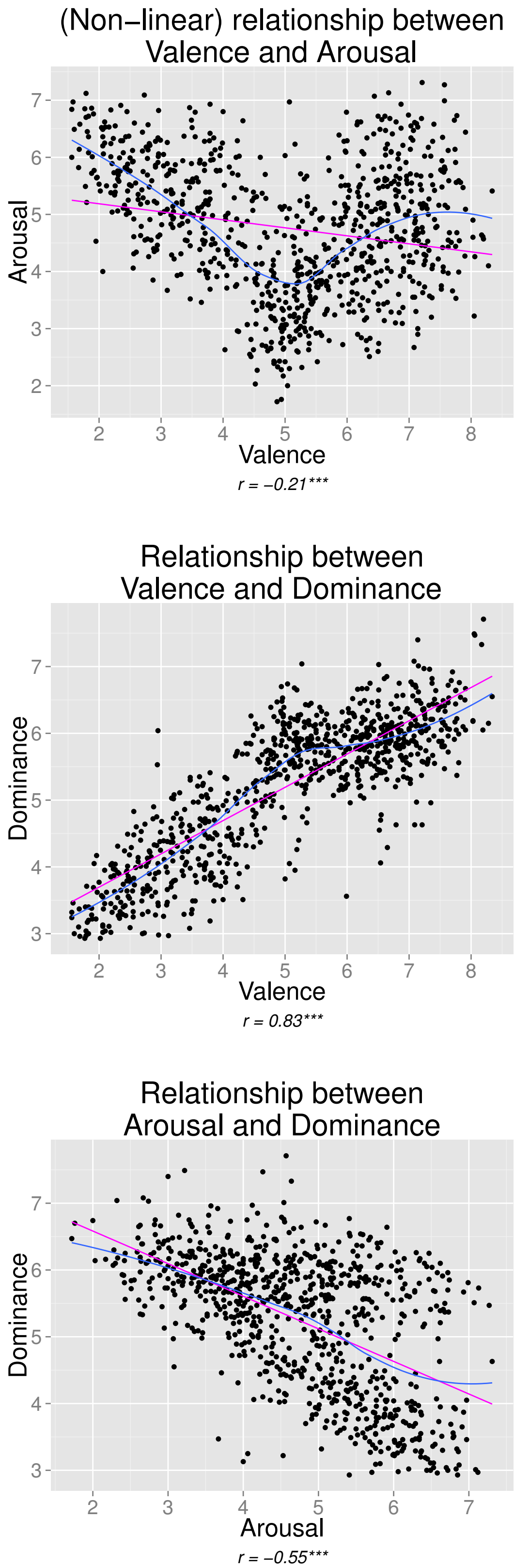


Figure 2.1: Correlations between the Pleasure/Valence Arousal, and Dominance dimensions, with deviations from linearity that give rise to the specific shapes of the relationships

2.2.3 Missing values

In terms of missing values, only the Valence and Arousal dimensions contained complete data. However, of the two Dominance distributions ("Dom1" and "Dom2")[2] included in the database, depending on which SAM rating scale was used in the measurement (Lang et al., 2008), "Dom2" contained considerably more missing data than "Dom1". Thus, we retained only the "Dom1" scale for further use[3], to benefit from its more complete data. We reduced the overall dataset accordingly, leading to a sample size of $N = 942$.

2.3 Results

2.3.1 Preliminary analyses

2.3.1.1 Outliers

Given the variety of emotional material included within the IAPS database, we employed a form of outlier identification as an objective means to filter out images exceeding the emotional intensity of stimuli expected in daily life, which could prove overly stressful for participants.

Outliers might also distort the clustering solutions (e.g., for k-means and model-based approaches), thus constituting an additional reason to identify and remove them. Specifically, outliers used with the model-based clustering might lead to a different number of clusters and/or alter the cluster memberships, without necessarily nesting outliers into a cluster of their own (Fraley & Raftery, 2002; Hautamäki, Cherednichenko, Kärkkäinen, Kinnunen, & Fränti, 2005; Wu, 2012; Xu & Wunsch, 2009).

Using the `R` language (R Core Team, 2015)[4], all three univariate distributions were found to be non-normal according to the Shapiro–Wilk test, so any method of determining outliers that was based on averages would probably be inappropriate (since the averages would not adequately represent the distribution). Hence, we opted for a more robust indicator: the Median Absolute Deviation (MAD; Leys, Ley, Klein, Bernard, & Licata, 2013)[5]. Therefore, images that were more than 2.5 MADs away from the median, in either direction, were removed before further analyses could be conducted. No outliers could be identified using this method in the Valence or Arousal distributions,

[2] "Dom1" refers to the "classic" SAM Dominance scale, whereas "Dom2" refers to a version of the Dominance scale on which the SAM icon with the highest control presented a more assertive and dominant facial expression/posture than in the classic version.

[3] The correlation between "Dom1" and "Dom2" was remarkably high ($r = .98$), for the $N = 60$ cases measured on both versions of the Dominance scale. Thus, we were able to safely use only "Dom1" in our analyses.

[4] The `R` code for our analysis is available at `www.github.com/CaterinaC/IAPSClustering2016`.

[5] According to this method, acceptable values should lie between the median $\pm$ (x * MAD), where we opted for $x = 2.5$.

but interestingly, 32 images[6] were flagged as outliers due to their Dominance scores, and were thus removed. This was done to avoid distorting the clustering solutions subsequently, and also to filter out potentially harmful material, in an empirically principled, replicable manner.

2.3.1.2 Representativeness/precision of measures

Additionally, we implemented a measure to ensure the precision of the stimuli to be used: building 95% confidence intervals (CIs) around the normative image ratings, to give an indication of how precisely the population means could be estimated, on the basis of the sample averages from the approximately 100 participants rating each image. We selected stimuli with CIs spanning no more than one point in total around the normative rating, which we considered to be sufficiently narrow, given that the three dimensions were measured on 9-point Likert scales. Using this criterion, 61 cases that were judged too imprecise were removed, since they could subsequently affect the inferences in our study; 46 images were removed due to the width of their CI on one dimension, 13 due to their CI width on two dimensions, and finally, two cases with CIs too wide on all three PAD dimensions simultaneously. After we had removed cases on the basis of both outlying values and CI widths, the sample size was reduced to $N = 849$.

2.3.2 Clustering techniques

After employing the filtration methods described above, three clustering procedures - k-means, hierarchical, and model-based clustering - were used to produce a set of coherent clusters that could be used in later primary research. For the reasons explained previously, the clusters were built on the basis of the normative ratings for all three available measures associated with the IAPS: Valence Arousal, and Dominance.

2.3.2.1 K-means clustering

When using this method, various indices were consulted to identify what the appropriate number of clusters (k) should be, including the Caliński–Harabasz Index (Caliński & Harabasz, 1974), the Ball Index (Ball & Hall, 1965), and the Hartigan Index (Hartigan, 1975), which are all based on within-/between-cluster sums-of-squares calculations (i.e., minimising the former and/or maximising the latter to ensure cluster compactness and/or the separation between clusters), as well as the Simple Structure Index (SSI; Dimitriadou, Dolničar, & Weingessel, 2002; Dolnicar, Grabler, & Mazanec, 1999), and

[6] IAPS codes: `3000, 3001, 3010, 3015, 3053, 3059, 3063, 3064, 3080, 3102, 3131, 3170, 3266, 3500, 3530, 6230, 6231, 6250, 6250.1, 6260, 6263, 6300, 6350, 6510, 6520, 9075, 9252, 9410, 9413, 9600, 9908, and 9940`.

others. The general trends shown by some of these indices are presented in Figure 2.2, where the nature of the dataset is such that various clustering indices detect different characteristics of the data and do not converge on any simple answer as to the "correct" number of clusters that should be extracted. For further details on these and other indices, please see the supplementary material (Appendix H.2, p.562).

On the one hand, it may seem surprising that a subset of over 800 IAPS images may have several k-means clustering criteria peak for the number of only two[7] or three clusters, considering the amount of variation in both the content and scores of the IAPS images. However, this could be accounted for theoretically by the emergence of a dichotomous "Positive and Negative Affect" structure (PA/NA, developed more in the Discussion), sometimes accompanied by the natural emergence of an additional neutral cluster. In Figure 2.3, both clustering solutions are displayed using colour coding for each cluster in the 3-D space, and are shown to cover extensive areas of the 3-D space.

On the other hand, higher values for k might be more suitable for the data, as is suggested in Figure 2.4, which shows that as the number of clusters increases, so does the amount of explained dissimilarity between the cases (calculated as 1 − *unexplained dissimilarity*, or 1 − *within-cluster dissimilarity*). Thus, as the number of clusters increases, within-cluster homogeneity also increases. However, k-means does not penalise for the increasing number of clusters (unlike model-based clustering), so that, conceivably, the total amount of dissimilarity would only be explained when the number of clusters equalled the number of cases. In other words, there is no single, definitive cutoff to determine which value of k best fits the data.

Since there may be arguments against using either a very small (e.g., $k = 2$ or even $k = 3$, with too many heterogeneous cases blended in the same group, as shown in Figure 2.3) or a very large number of emotional categories (e.g., $k \geq 8$, leading to a very fragmented and unparsimonious structure, with relatively few cases per cluster), we now turn to the other clustering methods for additional solutions.

2.3.2.2 Hierarchical clustering

Jointly testing various linkage methods (i.e., strategies for progressively merging clusters, described in more detail in the supplementary material - Appendix H.2) and distance metrics allowed us to find the combination yielding the clustering solution with the highest degree of similarity to the original data (or matrix containing the distances between every pair of IAPS cases). We found that Average Linkage (i.e., merging clusters based on the average distance between their points) paired with correlation-based distances (i.e., assigning cases to clusters on the basis of correlations) produced the results most

[7] Please refer to the supplementary material in Appendix H.2 for more details on the measures of Connectivity and Average Silhouette Width that suggested this value.

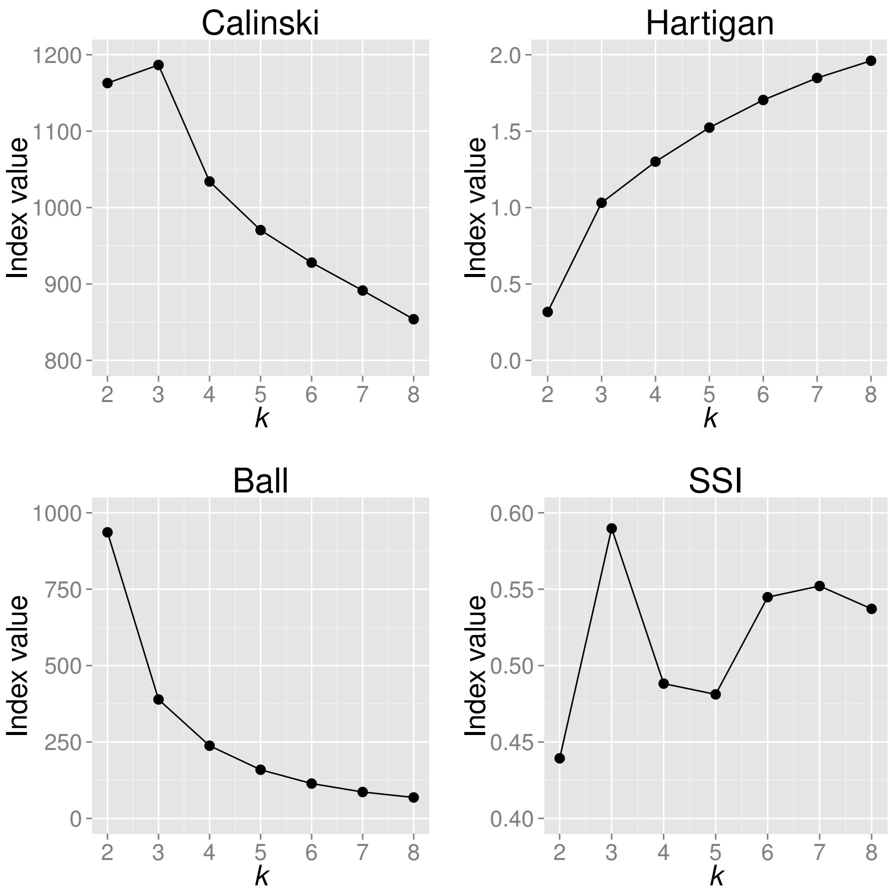


Figure 2.2: Various clustering indices indicate different "optimal" values for k. This graph may change slightly with every run of the clustering algorithm, due to the random seeds k-means uses. As such, 100,000 repetitions were run on the k-means clustering algorithm, each time with a range for k from 2 to 8, and with the values of the Calinski, Ball, Hartigan and SSI criteria computed each time (with the Ball criterion having to be minimised, unlike the other three criteria, which must be maximised). The average values for these criteria were then computed across all the repetitions and indicated (left to right, and top to bottom) that 3, 8, 8, and 3 clusters should be extracted, respectively.

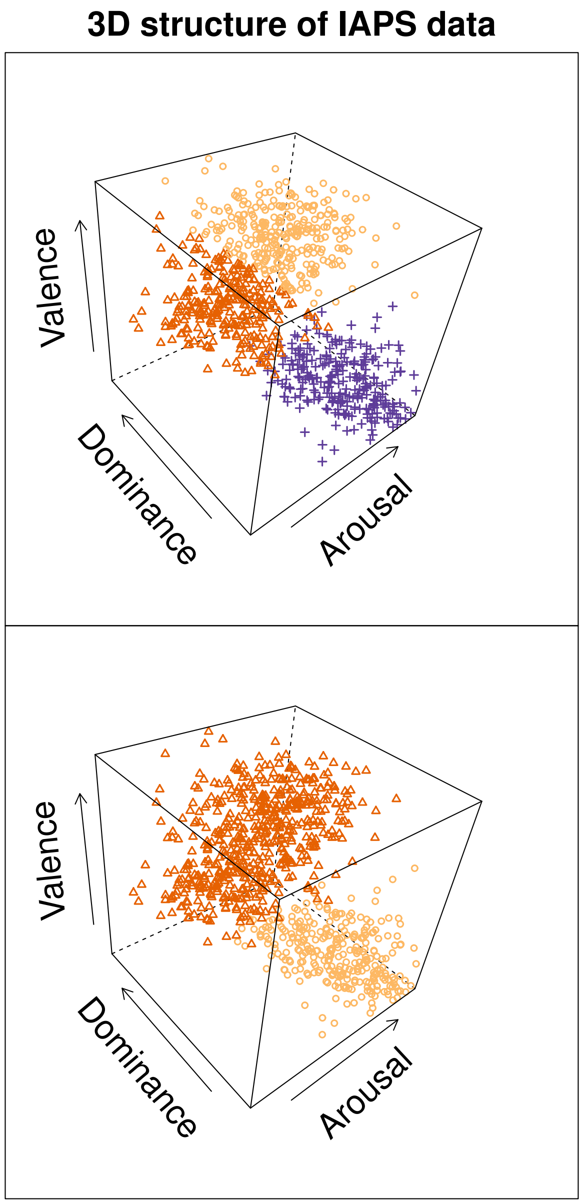


Figure 2.3: Data structure of IAPS. It is worth noting that large portions of the 3D space remain unpopulated, signalling either that the IAPS does not cover those combinations between Valence, Arousal and Dominance, or that photographic material in general would have difficulty with this.

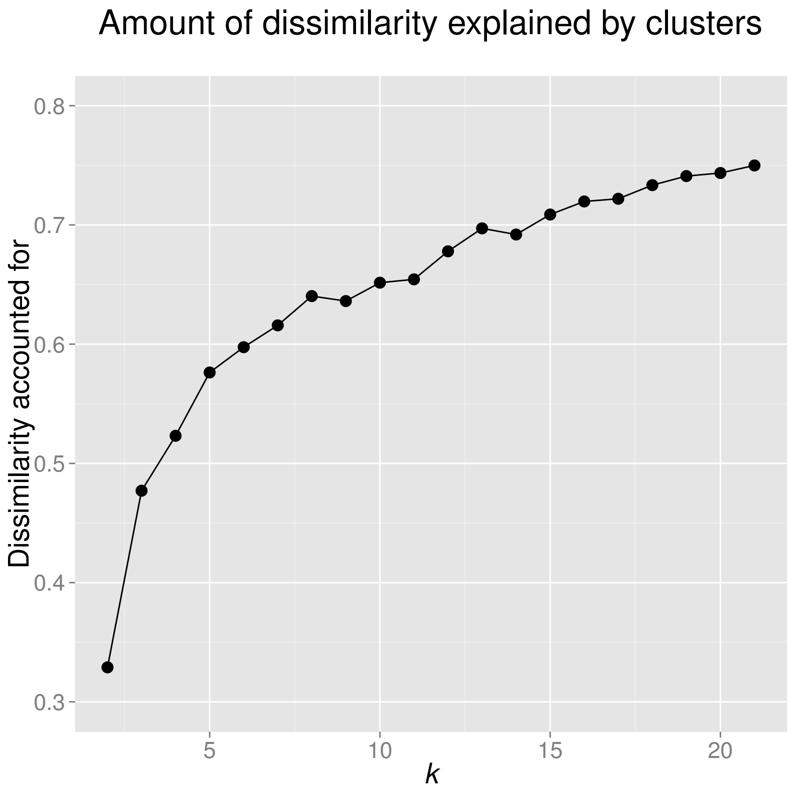


Figure 2.4: The amount of dissimilarity (as computed using `R` package `clue`, Hornik, 2005) between cases is accounted for by ever-increasing values for k.

similar to the original distance matrix (cophenetic correlation, r = .91). Consequently, this combination was the most suitable for the IAPS data, and shows how essential PAD relationships are when determining how to group the IAPS images. The next best result was attained by Single Linkage (in which cluster-merging depends on the distance between the closest points belonging to different clusters), again combined with correlation distances (r = .87). Thus, after having reconfirmed the importance of the PAD correlations and identified the most suitable hierarchical agglomeration method for this dataset, we proceeded to determine the most appropriate number of clusters in the data.

In terms of connectivity, average silhouette widths, and Mantel optimality (briefly described within the supplementary material, in Appendix H.2), a number of two clusters was suggested, whereas the Dunn Index indicated three. This corroborates the findings from some of the k-means indicators, and suggests the overall strength of the PA/NA structure within the IAPS, with or without an additional neutral cluster. However, as with k-means, some variability was to be found; for example, when using the elbow method for partitioning variance into clusters (using the `GMD` `R` package; Zhao & Sandelin, 2012), the optimal number of clusters (also based on average linkage) indicated was seven. Other clustering indices suggested nine clusters; however, still others provided more discrepant results, indicating numbers ranging from four to 15, or as many as 30 clusters. Overall, the most endorsed options were two (perhaps three), or nine clusters. For more information, please see the supplementary material attached in Appendix H.2.

2.3.2.3 Model-based clustering

Model-based clustering yielded a mixture model containing five clusters of varying Volumes, Equal (ellipsoidal) shapes, and Varying orientations (VEV), which is presented in Appendix A (p. 429). This model/configuration was optimal in terms of BIC values: BIC = –6,341.11, relative to the global minimum BIC value[8] for other cluster numbers and configurations, BIC = –8,671.93 (for one spherical cluster, with either equal or variable volume, and the configurations abbreviated as EII and VII, respectively). The second best BIC value achieved was –6,343.72, for a VEV model with four components (clusters). Full details regarding the BIC values for all the models considered can be found in the supplementary material (Appendix H.2).

The five-cluster solution proposed by the algorithm is described in Table 2.2, in terms of cluster centroids, sample sizes, mixing proportions (i.e., proportion of the mixture/overall sample that has been assigned to each cluster), and average uncertainties. By-cluster boxplots are also displayed in Figure 2.5, comparing the relative spreads of the clusters' Valence Arousal, and Dominance univariate distributions. In addition, given

[8] `Mclust()` in `R` seeks to maximise BIC values, given that it uses the negative of deviance.

the cluster centroids presented in Table 2.2, it is apparent that this clustering solution presents a symmetrical format: two negative clusters (one more so than the other), one neutral cluster, and two positive clusters (one more so than the other).

Finally, we assessed whether the assumption of multivariate normality held for these clusters, and found that, overall, the clusters presented ellipsoid shapes consistent with this assumption, with some further evidence also added by various multivariate normality tests. Please see the supplementary material (Appendix H.2) for details on testing the assumptions required for model-based clustering.

Table 2.2: IAPS cluster centroids, cluster sample sizes, mixing proportions (or percentage of total cases assigned to cluster), and average uncertainties extracted using model-based clustering.

Cluster	Valence	Arousal	Dominance	N	Mixing prop.	Average uncertainty
1	3.56	5.179	4.342	244	0.287	0.086
2	7.27	4.692	5.959	71	0.084	0.258
3	2.27	5.872	3.545	71	0.084	0.239
4	5.048	3.305	5.836	152	0.179	0.169
5	6.444	4.819	5.902	311	0.366	0.138

2.3.3 Validating the clustering solutions

After having employed three candidate methods - k-means, hierarchical, and model-based clustering - we proceeded to compare them on the basis of various validation techniques (full details are in the Appendix H.2), to select just one for further use. Given that variations were observed in terms of the “optimal” number of clusters suggested for k-means and hierarchical clustering by each clustering index, we deemed it appropriate to emphasise and pursue model-based clustering, which proved less affected by these issues, and also provided more information about the classification in the form of membership uncertainties. For a more meaningful comparison between the methods, parsimonious clustering solutions were formed using each of the three algorithms for a number of $k = 5$ clusters, as was suggested by model-based clustering.

2.3.3.1 Finding a stable structure within the data, across methods

Assuming that the IAPS data present a clear, discernible structure, all of the clustering algorithms should in principle be able to identify this structure despite their computational differences. To check this, we assessed the extent to which model-based clustering yields membership assignments that overlap with those from the other two competing methods.

The Variation of Information criterion (VI; Meilă, 2007) suggests that not much information is to be gained/lost when moving from one classification to another (i.e.,

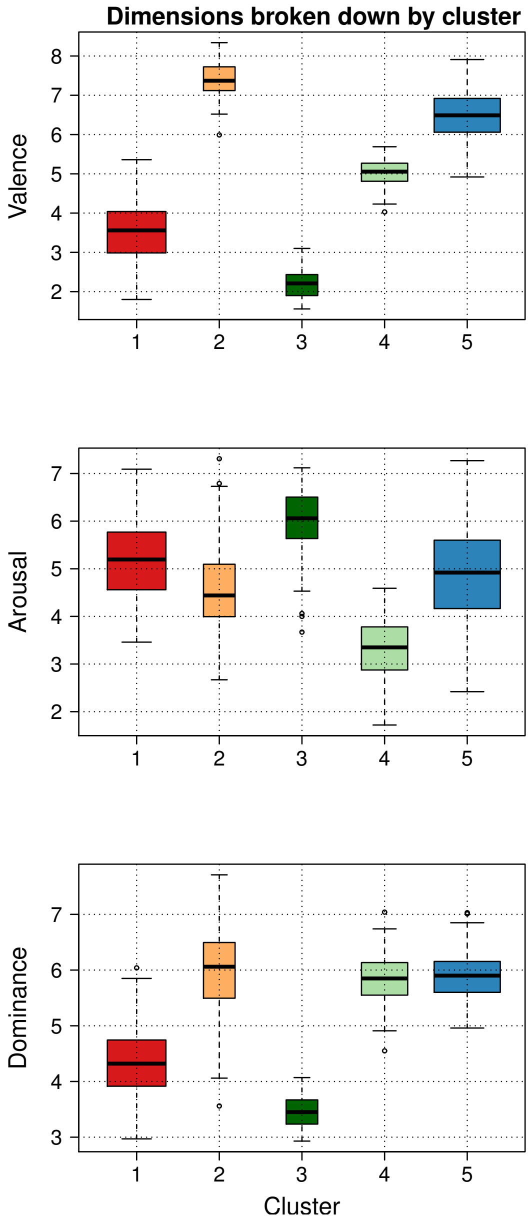


Figure 2.5: Cluster boxplots, for each dimension. The boxplots indicate, for each cluster (coded in colours), the spread of cases assigned to it, in terms of Valence, Arousal and Dominance. Boxplot widths are proportional to cluster sample sizes.

there is considerable similarity between partitions of five clusters, regardless of the algorithm used to produce them), with the normalised VI between model-based and k-means clustering = .176 and the VI between model-based and hierarchical clustering = .217 (please see Appendix H.2 for details). This finding was corroborated by the relatively strong association found between partitions using Cramer's φ (between the k-means and model-based classifications, $\varphi = .704$, and between the model-based and hierarchical classifications, $\varphi = .516$). Therefore, on the basis of the VI and Cramer's φ, there is considerable similarity between the five-cluster solutions provided by the different algorithms. However, for further results, including those based on the Adjusted Rand Index (ARI; Hubert & Arabie, 1985), please refer to Appendix H.2. Thus, on the whole these results constitute moderate evidence that a specific data structure can be identified in the IAPS, given the level of agreement between the clustering methods.

2.3.3.2 Evaluating the stability of the model-based clustering solution

We assessed the stability of the clustering solutions using various criteria, including split-half validation (i.e., dividing the IAPS data into two random halves and computing the level of association between the partitions created independently on these halves of the data) and jackknife validation (i.e., removing 10% of the IAPS data randomly across a few thousand repetitions and assessing changes in the structure of the clustering solutions). Overall, model-based clustering performed well, with a high degree of association present between how the random halves of the data were clustered, suggesting that the stimulus groups identified were well-supported. In terms of stability after the random removal of 10% of the data points, model-based clustering also outperformed both k-means and hierarchical clustering, for which typically only one cluster was then identifiable in the data (i.e., no grouping of the data points could be achieved after the removal of data points using these methods). For more details on these and further analyses, please refer to the Appendix H.2.

2.3.3.3 Selecting equal numbers of cases from each cluster

Given that the five clusters provided by model-based clustering differed in size, a procedure was required to sample equal numbers of cases from each cluster that would represent their respective cluster to the highest degree. Since levels of certainty are also provided for each image during the model-based clustering process, these could be used to create a hierarchy in terms of how likely it was for each image to belong to the cluster it was assigned to.

Consequently, a given number of images could be selected according to their rank in this hierarchy (i.e., the first n most likely cluster members). Figure 2.6 shows the default distinction made by `Mclust()`: Cases with uncertainties below the 75^{th} percentile are considered acceptable, uncertainties between the 75^{th} and 95^{th} percentiles are risky candidates, and those over the 95^{th} percentile should not be used, as they do not show clear membership to a given cluster. We made the same distinction in our final results[9], where we indicate which IAPS images were assigned to which cluster, as well as the level of uncertainty associated with this classification - particularly, which uncertainties were above or below the 75^th percentile (i.e., whether or not they should be sampled for research). These results are suitable for researchers to use in most research contexts.

In our example, only the first 20 cases in the hierarchy of uncertainties were retained for closer inspection. These can be judged as the best representatives for each given cluster, and are portrayed in Figure 2.7, with the first five of each cluster also displayed in Figure 2.8, where they are shown to be meaningfully related to one another.

2.3.3.4 A comparison between our method and ad-hoc approaches to selecting IAPS stimuli

Studies relying on more typical, ad-hoc methods for sampling IAPS stimuli may face several risks. On the basis of a Google Scholar search for "IAPS images", we selected a small number of studies randomly from several pages of results. However, we only retained articles that also specified the IAPS image codes used, rather than simply the average PAD values for the images selected. We then assessed how the categories used in these studies matched our own.

First, as is shown in Table 2.3, the images intended to represent different affective categories in these studies sometimes share the same clusters that our model-based clustering uncovered. For instance, in the Glenn, Blumenthal, Klonsky, and Hajcak (2011) study, four images considered neutral and ten images considered pleasant all belong to one of our *positive* clusters (i.e., Cluster 5; see also Table 2.2 for cluster descriptions).

Second, the negative or positive stimulus groups used in studies tend to pool together stimuli that our method has distinguished as reflecting two types of positive or of negative material. For example, the Koenigsberg et al. (2010) study used a group of stimuli wholly considered to be negative; however, our method divided these between two separate clusters - one that is mildly negative and moderately arousing, and one that is more negative and more arousing, and with lower Dominance than the former cluster.

In some cases, a single stimulus category (i.e., neutral, on the basis of the research reviewed in Table 2.3) may spread across three or four of our clusters. For instance, in the study by Most, Chun, Widders, and Zald (2005), the neutral category in fact included 8 mildly negative images, 27 neutral images, and 20 mildly positive images,

[9] Available online for download in the repository at: `www.github.com/CaterinaC/IAPSClustering2016`.

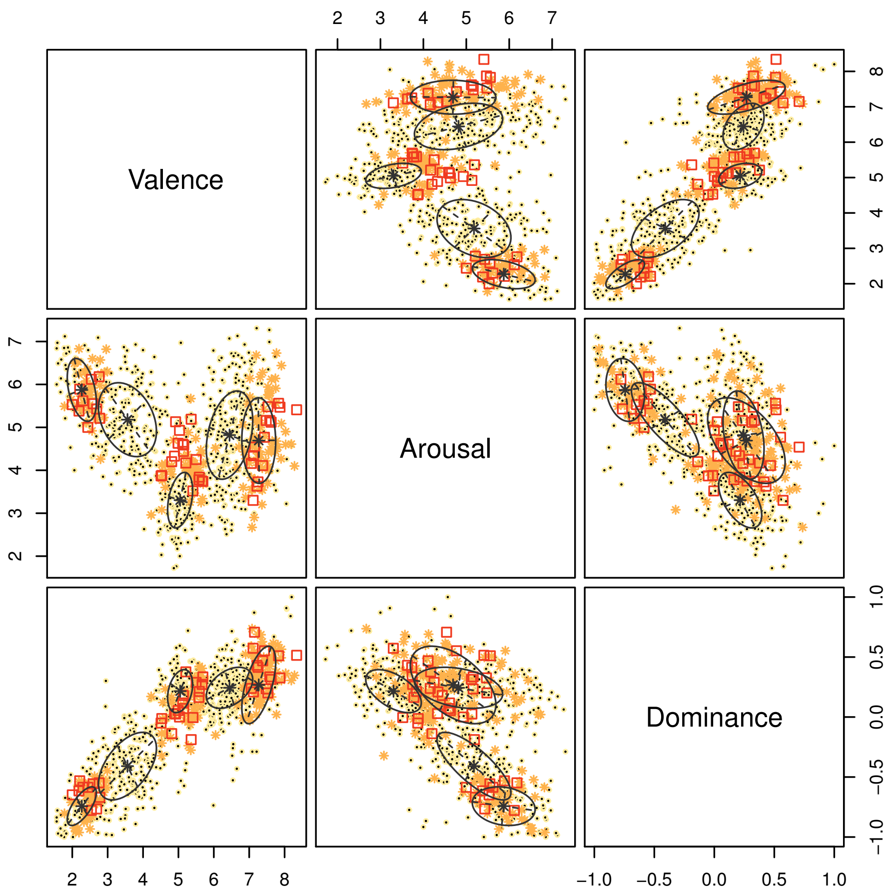


Figure 2.6: Bivariate scatterplots showing the default classifications of cases and the uncertainties provided by `Mclust()` in `R`. The uncertainties are coded using one of three symbols: ringed black dots for candidates with a high certainty of cluster membership; orange (light grey) asterisks for less clear cluster memberships; and red (dark grey) squares for cases to avoid using as stimuli, with very unclear memberships. Point size is an additional indicator for the level of classification uncertainty, with larger points indicating higher uncertainty.

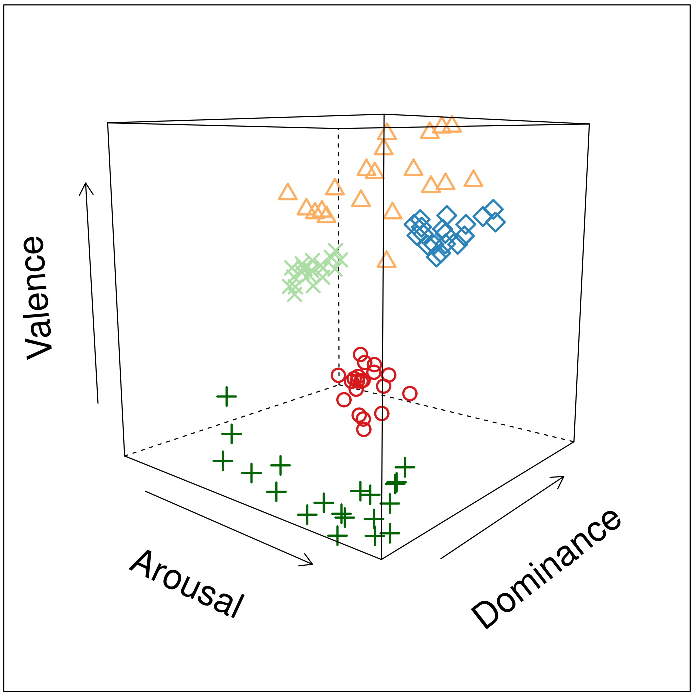


Figure 2.7: Selection of 20 most likely IAPS cases per cluster, for the $k = 5$ clustering solution. The colour coding was used consistently to match Figure 2.8 below.

according to our method. Another example is the study by Mikels, Larkin, Reuter-Lorenz, and Carstensen (2005), in which a category of neutral images intended to differ only in brightness actually belonged to four different emotional clusters within our own classification.

In addition, from Table 2.3, it is also apparent that without filtering images on the basis of 95% CIs, less reliable image stimuli can be included in studies. For instance, in the case of the Stins and Beek (2007) study, seven less reliable (in terms of confidence interval widths) images were included in the group of erotic stimuli. Similarly, images have also been selected without taking into consideration Dominance - including some images we excluded precisely because their norms for Dominance were missing. Finally, IAPS data outliers have also been included in studies, which could pose some ethical risks, due to their emotional intensity, and warrant closer inspection.

2.4 Discussion

A variety of research areas rely on stimulus databases for experimental use. The IAPS is one such widely used database, having currently amassed approximately 3300 citations in Google Scholar (April, 2016). Yet, despite its extensive use, a standard stimulus

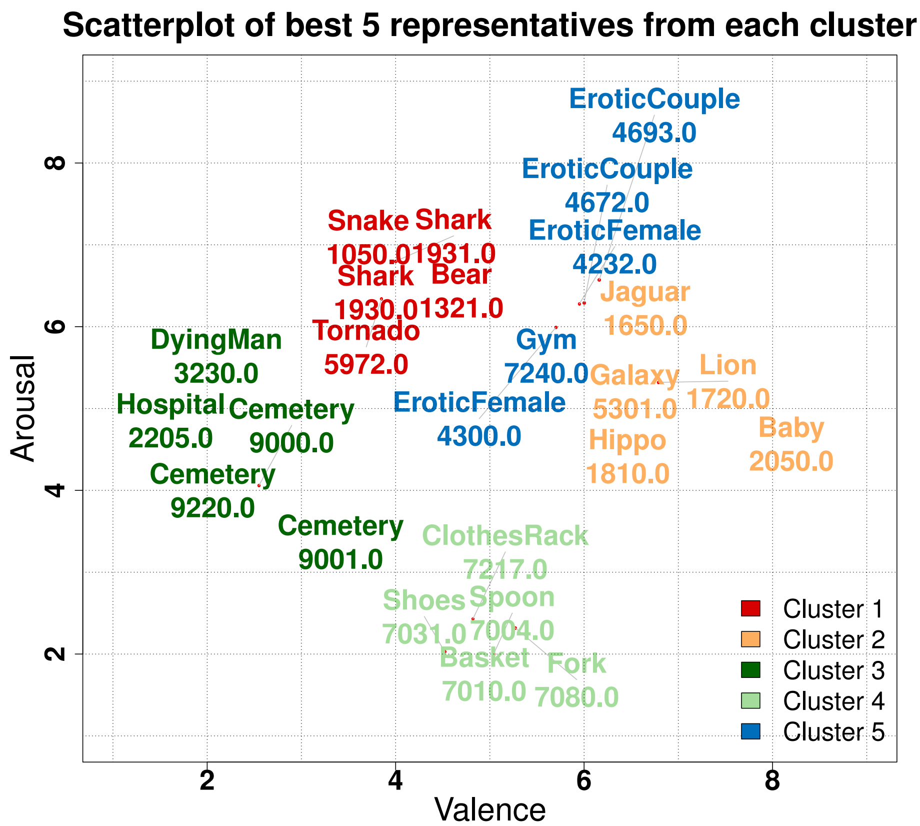


Figure 2.8: Selection of 5 most likely IAPS cases per cluster, for the $k = 5$ clustering solution, along with IAPS image codes. Colour coding was used consistently to match Figure 2.7 above.

Table 2.3: Stimulus groups used in various studies, redistributed according to our method.

Study	Stimulus categories	Total no. stimuli used per category	Image codes unaccounted for	Stimulus redistribution according to our method, and according to:							
				Missing Dominance	Outliers excluded	Wide CI excluded	C1 (−)	C2 (++)	C3 (− −)	C4 (±)	C5 (+)
Glenn et al. (2011)	Pleasant	18	–	–	–	3	–	5	–	–	10
	Neutral	18	–	1	–	–	–	–	–	13	4
	Unpleasant	18	–	–	6	3	4	–	5	–	–
Mikels, Larkin, et al. (2005)	Pleasant	80	17	–	–	10	–	22	–	–	31
	Neutral	80	8	–	–	–	7	5	–	41	19
	Unpleasant	80	–	–	3	2	71	–	4	–	–
Most et al. (2005)	Neutral	55	–	–	–	–	8	–	–	27	20
	Negative	39	–	–	11	6	5	–	17	–	–
Stins and Beek (2007)	Neutral faces	14	–	1	–	–	2	–	–	10	1
	Neutral household items	15	–	1	–	–	–	–	–	14	–
	Erotic	17	–	1	–	7	–	–	–	–	9
	Family scenes	12	–	1	–	2	–	8	–	–	1
	Mutilation	11	–	–	6	–	–	–	5	–	–
	Fear	18	–	1	5	1	9	–	2	–	–
Koenigsberg et al. (2010)	Negative	47	–	6	7	4	13	–	17	–	–
	Neutral	49	–	10	–	–	4	–	–	25	10

The clusters from one through five are represented by C1, C2, C3, C4, and C5, and refer to those described in Table 2.2. For the same columns, between parentheses we have included concise information about the Valence of our clusters, ranging from very positive (++) to very negative (− −). This has been done merely to aid interpretation of this table; however, as we previously stated, all three of the PAD dimensions were important in determining how the IAPS stimuli were assigned to these clusters, not only Valence.

selection strategy from the IAPS has yet to be devised - one that can easily take into account all three PAD dimensions simultaneously, and provide a stimulus grouping that is both empirically principled and optimal in terms of various statistical measures.

In this article, we proposed such a method based on the following sequence of steps: filtering out stimuli that constitute outliers or duplicates, and those with CIs wider than a preset criterion; creating stimulus categories using different clustering algorithms; and finally, validating these categories against several measures. Within the procedure we propose, we placed special emphasis on model-based clustering, an inferential method that provides not only a classification of the stimuli, but also an uncertainty estimate for each stimulus assigned to a cluster. Examining these uncertainty estimates allows researchers to control for how well stimuli reflect their underlying category and to select only those stimuli that reflect their cluster in the most meaningful way.

2.4.1 Filtering out stimuli prior to clustering

As a first step toward creating a selection of stimuli for experimental use, the MAD has proved to be a useful tool for identifying stimuli that may be ethically questionable, due to their violent or threatening nature. In addition, Grühn and Scheibe (2008) found that IAPS ratings for negative images tend to get more extreme with age. Thus, as a precautionary measure, filtering out outliers using the MAD might have to be considered more carefully depending on what sample/ population the stimuli are aimed at, as the same IAPS image might be more distressing for one category of participants than another.

Using the MAD, we were able to exclude 32 images due to their particularly low Dominance scores (i.e., in the case of highly violent images, with an average Valence level of 1.98 - e.g., image `3001`, a headless body; `3131`, mutilation; `3170`, a baby with a tumor; etc.). Interestingly, these same cases were not flagged as outliers given their scores on the other dimensions. This provides further evidence that Dominance scores reflect a different process of emotional evaluation and should be considered more frequently when selecting IAPS images. Relatedly, Dominance is believed to be more easily distinguishable from the other two dimensions in social situations (rather than photographic material; Bradley & Lang, 2007b, p. 32), further supporting its general inclusion in stimulus selection procedures, as an additional contributor to emotional experiences.

The large standard deviations associated with the ratings for most stimuli from the IAPS have usually resulted in wide 95% CIs (spanning more than one point on the nine-point Likert scale used for ratings). However, within our overall approach based on CIs, other (more or less conservative) criteria may also be applied regarding the width of these CIs, depending on researchers' specific aims. This type of verification has proven to be highly useful either for deciding which stimuli to retain for the subsequent clustering procedure, and for a better appreciation of the amount of variability in the individual IAPS ratings leading to the normed means. Although we are unable to give an exact reason why some of the stimulus norms were insufficiently precise, on the basis of our criterion, these results clearly suggest a verification as simple as this should become a more standard practice when selecting stimuli from stimulus databases.

We would stress that it is possible for any emotional stimuli database to present these same concerns. This is because emotional stimuli are conceivably very subjective, thus leading to the large standard deviations observed, and implicitly, the lower degree of certainty as to how they may be perceived by individual participants (e.g., image `EroticFemale 4210` registered the highest standard deviation of all IAPS images, suggesting that reactions to it varied considerably). On the other hand, it is also possible these characteristics might be specifically related to the features of IAPS, but not of other emotional stimuli collections; thus, image quality and historic context, ecological validity, and so forth, may also be involved. Future work will be necessary to address this research question.

2.4.2 Clustering the stimuli

When using k-means and hierarchical clustering to classify IAPS images, the repartition of cases between clusters represents a separate step from choosing the "appropriate" number of clusters existing within the data. Our analysis showed that it is difficult to discern a clear cluster structure within the IAPS data. For example, in the case of k-means, the optimal value for k oscillated between two, three, or eight, depending on the clustering index used, and on the total number of clusters tested. Similarly, for hierarchical clustering, a number of two, three, seven, or nine clusters was indicated as suitable for the IAPS data, also depending on the index and number of clusters. It may seem surprising that a number of clusters as low as two, or even three, could be suggested by both k-means and hierarchical clustering, for a sample size as large as $N = 849$ images, varying considerably in terms of Valence Arousal, and Dominance scores. However, the emergence of these solutions is understandable, for theoretical reasons and/or due to the shape of the IAPS data.

First, the $k = 2$ solution carries theoretical significance by corroborating principles used in the construction of the Positive and Negative Affect Schedule (PANAS; Watson, Clark, & Tellegen, 1988), since the two emerging clusters can be interpreted as matching the Positive and Negative Affect components of the scale, which measure the corresponding affective moods with adequate reliability and validity. This similarity directly indicates that clustering methods can provide meaningful results, which can be validated against current practices and/or theory.

Second, the nonlinear ("U" shaped) relationship between Valence and Arousal can easily be split into three sectors, a characteristic that carries over into 3-D space, when Dominance is added. Thus, one cluster is negative with higher Arousal, another is neutral with lower Arousal, and the third is positive, again with higher Arousal. Although this three-cluster solution may appear similar to those from typical image selection practices (cutoff points and/or factorial designs, centred on selecting three Valence groups: negative, neutral, and positive), it differs from these approaches in that it accommodates all three PAD dimensions simultaneously with ease, and also takes the structure of the data into account, without imposing unsustainable assumptions (i.e., independence of the PAD dimensions). In fact, even if hierarchical clustering did not provide the final classification of the IAPS data, it did reveal most clearly the importance of the PAD relationships, since using correlation-based distances always yielded the highest correlations with the original data for this clustering method. This suggests that the PAD correlations should always be taken into account when selecting stimuli from the IAPS, whereas using factorial designs without concern for them may simply lead to inappropriate groupings of stimuli, and subsequent experimental results that are difficult to interpret.

However, both of these solutions ($k = 2$ and $k = 3$) focus on the creation of just a few, large clusters, which would thus cover considerable portions of the 3-D affective space within the PAD model. As such, one large negative cluster would, for instance, include images with both moderate and higher Arousal, or both moderate and lower Dominance - leading to a lower degree of experimental control.

On the other hand, from a practical standpoint, the larger numbers of clusters (seven, eight, or nine) indicated by k-means and hierarchical clustering may be as intractable as the lower numbers, but for different reasons. Rather than blending together too many heterogeneous cases, when using a larger number of small clusters - the more clusters are extracted, the closer their centroids necessarily become and thus their "best representatives" are also drawn nearer. This can result in a potential reduction in statistical power. Also, more clusters (or treatment levels) would generally signify longer testing times and study expenses, which is not always feasible. Finally, smaller cluster sizes would be less useful for experiments requiring larger numbers of stimuli of the same type (i.e., from the same cluster).

In contrast to the previous two methods, model-based clustering uses a soft clustering approach,which provides an estimate for the degree of cluster membership (uncertainty) associated with each image. This allows for finer-grained control over stimuli used in experiments, which in turn can help make research inferences stronger. This method also provides additional flexibility in terms of adaptively distinguishing a variety of cluster configurations, thus being capable of a closer fit to the original data. In contrast, k-means would, for instance, favour spherical clusters in particular (Jain, 2010). Finally, unlike for k-means or hierarchical clustering,the optimal number of clusters in model-based clustering is assessed using the BIC, which penalises for large numbers of clusters, and simplifies the process of choosing which number of clusters to extract from the data.

In our case, a number of five clusters was suggested, which also represents a good compromise from a practical standpoint. In addition, the clusters were determined to be of Varying volumes, Equal shapes (i.e., ellipsoidal, rather than spherical), and Varying orientations within the 3-D space. The cluster centroids also suggest that for participants, "neutral" images present medium levels only on the Valence scale, rather than in the whole PAD model, as might have been assumed. Thus, neutral IAPS images tend to be somewhat lower in Arousal and higher in Dominance: For instance, a picture of a mug (IAPS code `7035`) intuitively seems "neutral", but this translates into medium values only on the Valence dimension (norm = 4.98), whereas the lower Arousal (norm = 2.66) suggests a more calming influence, and the higher Dominance (norm = 6.39) suggests very unchallenging content.

Equally, we have shown that two forms of negative and positive material exist, rather than one of each, which is the typical grouping used in research. For instance, we found that very negative content (e.g., IAPS image `Mutilation 3030`) presents very low Valence (as expected) but, uniquely, higher Arousal and lower Dominance. Thus, collectively, these three components (and not just Valence) seem to form what is usually perceived as "very negative" content. A second, milder, type of negative content was identified, as well, which still presents Valence values below the scale midpoint, but less extreme Arousal and Dominance values (e.g., IAPS image `Cigarettes 9832`). Similarly, positive content can also be divided into two subtypes using our method: positive, more arousing content (e.g., IAPS image `Erotic Couple 4693`) and very positive, more serene/less arousing content (e.g., IAPS image `Nature 5220`) - with both of these categories being fairly similar in their mean-level Dominance.

This five-cluster option generally benefits from empirical support based on the methods we employed to verify this. We first noted a moderate overlap between how the images were classified into five groups by k-means, hierarchical, and/or model-based clustering, depending on the measure used to assess the overlap. Although no structure is unanimously accepted within the IAPS data, measures such as the Variation of Information (VI) or Cramer's φ both suggested that $k = 5$ is relatively well-supported, even if each clustering method can shed its own perspective on the data (i.e., the amount of overlap was not maximal, which we discuss in more detail in Appendix H.2).

Subsequently, to ensure that model-based clustering is indeed the most suitable algorithm for use with the IAPS data, we removed 10% of cases randomly across a few thousand repetitions (using jack-knife validation), each time assessing how the optimal number of clusters changed. Ideally, if a robust clustering solution was found using a certain clustering algorithm, the removal of 10% of the values should make little difference. In the case of k-means and hierarchical clustering, this frequently resulted in only one all-encompassing cluster being identified in the data, which was deemed inappropriate. In contrast, model-based clustering showed more stability, and most often suggested $k = 3$ (followed by $k = 4$) as the optimal solution in this case. However, cross-tabulations showed that these model-based solutions were very closely correlated to the $k = 5$ solution achieved on the full dataset, and did not present any deeply concerning changes such as the cluster structure collapsing entirely (i.e., when we found just one cluster using the other two methods). Therefore, the differences seen in the values of k most likely reflect the fact that one or two clusters from the $k = 5$ solution were collapsed due to the induced data attrition (–10%), but that similarities between the solutions nevertheless remained robust.

Finally, when predicting the clustering structure of a random 50% of values based on that of the other 50% (using split-half validation), and comparing this prediction to the observed model-based classification of the target half, the two matched very closely. On the basis of all these indicators, we concluded that the five-cluster mixture model is well-supported by the IAPS data.

2.4.3 Method summary and recommendations for use

As an outline for our method, we recommend first inspecting the IAPS images and filtering out duplicates, outliers, and images with CIs larger than a preset criterion (we opted for one point in total, on the Likert scales used for the IAPS norms, but researchers may be more conservative if they have specific reasons for this). Subsequently, on the basis of the findings detailed above and in Appendix H.2, we recommend resorting to a model-based clustering algorithm, which will nest the remaining images into five clusters, while also taking into account Arousal and Dominance in the creation of these clusters, even if researchers may only be explicitly interested in, for instance, Valence.

Regarding any more practical issues that may arise, we recommend maintaining this well-supported, five-cluster structure even if researchers may be interested in comparing fewer categories. For instance, assuming that a study is aiming to compare the effects of positive versus negative Valence on an outcome variable, just two of the five clusters may be used, which are farthest apart on this dimension, rather than altering the clustering solution to provide just two clusters in total.

Given that model-based clustering is a soft clustering method, cases were also assigned a level of certainty for belonging to their cluster. Unequal cluster sizes (with some of them being perhaps too large to be used in an experiment in their entirety) led to cases being sorted in descending order of their certainty of membership. This enabled us to select a constant number of images per cluster for subsequent use in an experiment -

those at the top of the hierarchy formed (i.e., with the highest certainty of membership, or equivalently, with the lowest uncertainty). Besides providing the ability to flexibly tailor this constant to the requirements of individual studies, these stimuli can also act as the best representatives of their respective clusters.

For illustrative purposes, five to 20 cases per cluster were sampled in the order of their certainty of belonging to their given cluster. This resulted in groups that are intuitively meaningful, with one very negative cluster including death-related scenes (e.g., hospital, cemetery, dying man); a second negative cluster including dangerous agents, which was higher in Dominance than the former one (e.g., snake, bear, shark); one neutral cluster that was low in Arousal and higher in Dominance (e.g., spoon, shoes, basket); one positive cluster including arousing scenes (e.g., erotic scenes, gym); and finally, another very positive cluster including less arousing "natural" scenes (e.g., hippo, jaguar, galaxy).

Depending on the number of stimuli required per cluster for individual studies, researchers may also wish to know how many stimuli can safely be sampled from the clusters, in their order of membership certainty. One solution could be to use the criteria from the default `Mclust()` (Fraley & Raftery, 2006) graphical output in `R`, which considers images with uncertainties below the 75^{th} percentile to be appropriately clustered. Of course, more conservative cutoffs could be selected, should the amount of data support it, the number of stimuli required be relatively small, or the study imply high stakes (e.g., in clinical research).

If, on the other hand, researchers require larger numbers of images per cluster than, for instance, those having uncertainties below the 75^{th} percentile, or even more than the size of the smallest clusters extracted (e.g., $N = 71$, in our case), several solutions exist. First, one can relax the reliance on uncertainties when excluding images, but nevertheless retain the uncertainties for use as statistical weights in models, after experimental data have been collected. This would ensure that better cluster representatives would count more when determining the research results, making images with higher uncertainties still usable. A second alternative could be to resort to sampling additional photographic stimuli from other databases. To the extent that PAD ratings/norms exist or can be obtained for such images, it would be trivial to determine their cluster memberships with regard to the present results.

Finally, it is also possible for researchers to modify our method to suit their aims - for instance, in terms of the criteria used for the CI widths, or the level of uncertainty used to determine clear cluster memberships - as long as there is good justification for doing so and deviating from the standard approach (e.g., in clinical research with high stakes).

2.4.4 A comparison between our method and ad-hoc approaches to selecting IAPS stimuli

On the basis of our brief comparison, we discovered that a common practice is to group together stimuli that, according to our method, actually represent different types of negative or positive images (e.g., when a single group of positive material is used, instead of one positive cluster of "serene scenes", with lower Arousal and somewhat higher Dominance, plus one cluster of "exciting scenes", with higher Arousal and somewhat lower Dominance). Thus, a single, generic grouping of "positive" (or "negative") images may obscure any specific effects due to just one *type of* positive (or negative) material - particularly if the effects actually differ between the several types of positive (or negative) images.

This would be in addition to the relatively frequent inclusion of outliers in the literature, and importantly, of less reliable images (with 95% CIs wider than one point). Of these, outliers could be ethically risky, and should be avoided especially when relying on cluster analysis for stimulus selection (otherwise, they may distort the clustering solutions), whereas images with wide CIs can introduce additional error variance into research results.

Another interesting finding that emerged from our comparison is that effects can become diluted if neutral categories are not truly neutral, and extend into the space of clusters that we have found to actually be mildly positive or negative. This could result in diminished power to detect differences between the "neutral" and positive or negative stimulus categories.

Finally, we would underline that we do not wish to highlight these differences as criticisms of previous research using the IAPS. Rather, it is our intention to improve on these very widespread methods for selecting stimuli, by promoting our novel method that relies on model-based cluster analysis. Indeed, we believe previous image selection techniques may still be useful in limited contexts; however, it would be very difficult to predict when or to what extent they might influence results (by obscuring effects or "diluting" them, etc.). In addition, they may often vary considerably from study to study (in terms of both selection criteria and resulting selections), making comparisons between studies more difficult. As such, we argue that relying on a statistical, easily reproducible[10] and automatic procedure, which also quantifies the extent to which images belong to a given cluster, is much to be preferred.

[10] As long as any researchers using model-based clustering are transparent about all of the settings/data-cleaning methods used with the procedure.

2.4.5 Further research and limitations

Despite being arguably more objective than "manual" selection methods, cluster analysis is not an "exact science." As has been shown previously, the large variety of algorithms available can lead to substantial variations in clustering solutions. It is sometimes partly up to the researcher to decide which clustering solution is appropriate for their data. This is particularly the case with k-means and hierarchical clustering, because the clustering process is initialised using random seeds and/or various clustering indices that may suggest conflicting numbers of clusters. In contrast, with model-based clustering such difficulties can largely be avoided, because the results are identical on different runs of the algorithm (unlike k-means), and the only relevant criterion for choosing the number of clusters is the BIC.

Thus, any flexibility attributed to clustering methods (model-based clustering, in particular) may be seen as an asset, rather than a risk for objectivity, as long as the choices made by researchers (i.e., level of uncertainty, the width of CIs, etc.) are transparent and justified by convincing arguments. The present work aims only to provide a guide for a method that is more appropriate than manual selection strategies - particularly if multiple dimensions are used simultaneously for selecting stimuli.

In addition, although the cases sampled from each cluster acquit themselves of being good cluster representatives, the overall selection of treatment levels (or clusters) is ultimately constrained by the type of data in the IAPS - or whichever stimulus database would be used in research. As such, the final selection of stimuli cannot include categories of stimuli that are not part of the database to begin with. In the case of IAPS data, this may be either because such stimuli would be difficult to find, due to the PAD correlations (e.g., very negative images with low Arousal are unlikely), or because the IAPS domain of images does not include emotional material that extends as far as possible within the 3-D PAD space (e.g., images with moderate Valence and moderate, rather than low, Arousal are not very common).

These concerns could be addressed in the future either by the inclusion of new images or by a renorming process for the IAPS database (potentially via Amazon Mechanical Turk), using larger samples to rate each image. This can also present the added benefit of the average values being more stable (i.e., smaller standard deviations), and therefore fewer images being filtered out of the clustering procedure, thus creating more comprehensive clusters. However, until then, when interpreting results based on the current IAPS norms, the empty areas in the PAD space will require careful consideration, since otherwise research conclusions may be biased.

In terms of future research, an interesting avenue would be to compare empirical results when using a manual image selection method, relative to our cluster-analysis-based classification. Also, there is room yet for further standardisation of the IAPS images - for example, in terms of their spatial frequency content (i.e., their level of detail or "coarseness"), which may interact with their affective processing (Delplanque, N'diaye, Scherer, & Grandjean, 2007). Cluster analysis could take such dimensions (as well as participant age, etc.) into account when creating experimental treatment levels, provided they have been converted to standard scores beforehand. Furthermore, depending on whether the raw data used to produce the IAPS normative ratings will be made available, the source of the large standard deviations could be explored further, to indicate improved selection strategies.

Finally, for any research requiring "emotionally ambiguous" stimuli, which do not clearly fit into any particular cluster, uncertainty estimates for the classification of images may provide a more empirically principled means to identify these along multiple dimensions. This would represent a higher level of rigour, the application of which could be explored in future research.

2.5 Conclusions

In this article, we have presented a method for selecting experimental stimuli, which we have illustrated using the IAPS database. Using model-based clustering and Valence Arousal, and Dominance scores, we classified the IAPS images into five categories - with each image presenting a certain level of certainty of belonging to its respective cluster. Our method is flexible, efficient, and reproducible, and it provides meaningful clusters in a symmetrical format, in terms of their Valence ratings: two negative clusters (one more so than the other); one neutral cluster; and two positive clusters (one more so than the other). However, this method could easily be extended to other stimulus databases, in which the same principles may be applied: careful data inspection, including the removal of any duplicated cases in the stimulus database; the exclusion of missing values and outliers (in a judicious manner); selecting the most precise cases; selecting an appropriate clustering algorithm and clustering solution; and finally, extracting a constant number of stimulus exemplars from each cluster.

Chapter 3

Study 2: Selecting affective words and sounds to match IAPS images

3.1 Introduction

AFFECTIVE science aims to investigate how emotional processing unfolds, and/or how it impacts on other cognitive processes. In order to achieve this, researchers often resort to using one or multiple types of stimuli as test cases, to shed some light on how affective processes operate. However, the choice for which stimuli to use often goes unjustified - for reasons that are unclear. This might be due to an implicit assumption that the various stimulus types are largely equivalent, or it may be a matter of preference / practicality among scientists.

Nevertheless, we believe that comparing stimulus modalities and/or elicitation methods, and knowing beforehand how / if the stimulus type affects response patterns, is worth explicit scientific exploration, for three main reasons: firstly, this can make research results more interpretable; secondly, it can also simplify the process of choosing "appropriate" stimuli for research projects; and thirdly, it may explain some of the situations where findings fail to replicate.

In addition, any similarity or commonality found consistently between different stimulus modalities (e.g., auditory vs. visual etc.) could suggest the existence of an overarching system of emotions, in charge of supervising and handling the responses to these stimuli. However, this remains a fairly elusive endeavour, since an adequate test of this idea would require the inclusion of a variety of emotion elicitation methods (discussed in section 1.3, page 41), and also multiple types of dependent measures (see section 1.4, page 49 for details).

To achieve these goals, it is important to devise a framework allowing the exploration of different types of stimuli at the same time, while matching them as closely as possible in terms of any comparable dimensions, e.g., the PAD model, which includes Pleasure / Valence, Arousal and Dominance (Mehrabian, 1995, 1996; Russell & Mehrabian, 1977), or stimulus content. Pursuing this general area of research could therefore support a further, more practical purpose: matching stimuli on dimensions such as the PAD model can limit the influence of confounds in results.

Consequently, here we propose a systematic approach for selecting multi-modal stimuli from established databases, based on a set of reproducible statistical criteria, and for matching them as closely as possible on a few dimensions of interest, in order to make their subsequent empirical comparison more meaningful. We will illustrate our approach with images from the International Affective Picture System, or IAPS (Lang et al., 2008) discussed at length in our previous work (see chapter 2 in this thesis, and Constantinescu et al., 2016), the Affective Norms for English Words, or ANEW (Bradley & Lang, 1999a), and the International Affective Digitized Sounds, second version, or IADS-2 (Bradley & Lang, 2007b).

3.1.1 Affective images

The IAPS (Bradley & Lang, 2007c) represents a collection of several hundred images, each of which is associated with a normative rating on the PAD model, i.e., a norm for Valence / Pleasure, or how positive an item is; another for Arousal, or how alerting an item is; and finally, a norm for Dominance, or the amount of control the participant believes they have over the stimulus. These dimensions, together with relevant research examples, and strategies for sampling stimuli from the IAPS, were all discussed at length in Constantinescu et al. (2016), as well as Chapter 2.

3.1.2 Affective words

Verbal stimuli have been used in research over the past few decades for a variety of purposes, as a few examples will illustrate: Graves, Landis, and Goodglass (1981) for instance investigated whether there are hemispheric differences in between the genders, with regard to how emotional words are processed. Later, Kousta, Vinson, and Vigliocco (2009) studied whether emotional Valence lends any benefit for prioritising the processing of emotional words (positive or negative) over neutral words, in a lexical decision task. Kiehl et al. (2001); Williamson, Harpur, and Hare (1991) also used them as stimuli in research confirming that individuals scoring high in psychopathy tend to differ from controls in terms of how they process emotional words.

Verbal stimuli have also been organised into normed databases such as the ANEW, developed by Bradley and Lang), and made available in two versions: firstly, in 1999a, and then in 2010, when more words were added, but information about word frequency was removed. This is somewhat surprising, given that individuals can show some degree of sensitivity to word frequency, e.g., Kuchinke, Võ, Hofmann, and Jacobs (2007) found that pupillary responses during lexical decisions can vary with word frequency, but not emotional Valence. Regardless, the norms for both versions of this database were measured according to the PAD model, and similarly to the IAPS, they contain the average Valence (Pleasure), Arousal and Dominance score for each stimulus, as well as their standard deviations.

In order to allow for a wider usage of these stimuli, Stevenson, Mikels, and James (2007) conducted a study aiming to classify them according to any basic emotions they might reflect (i.e., happiness, anger, sadness, fear and disgust). Interestingly, their results suggest that these two forms of describing emotions (as discrete categories, or using continuous dimensions) are complementary, i.e., PAD values alone cannot perfectly account for how the emotional words are categorised by participants into discrete groups, and vice versa[1].

Given the influence of culture on verbal content generally, various alternatives to ANEW have appeared for non-English / non-native English speaking communities; among them: the Spanish adaptation of ANEW (Redondo, Fraga, Padrón, & Comesaña, 2007), The Berlin Affective Word List Reloaded (BAWL–R, Võ et al., 2009), the Discrete Emotion Norms for Nouns: Berlin Affective Word List (DENN-BAWL, Briesemeister, Kuchinke, & Jacobs, 2011), or the Nencki Affective Word List in Polish (NAWL, Riegel et al., 2015).

3.1.3 Affective sounds

Affective sounds have also been used in a variety of research contexts, with researchers investigating e.g., the physiological changes they may induce, their relationship to other cognitive processes, and their ability to influence self-report.

For instance, in terms of physiological responses, Partala and Surakka (2003) found that emotional sounds influence pupillary reflexes compared to neutral sounds, and Gomez and Danuser (2004) discovered that both negative and low-arousing sounds, as well as positive and highly-arousing sounds, are associated with faster breathing, and that overall Arousal levels are linked to increases in heart rate. In addition, findings by Bradley and Lang (2000) further suggest that electrodermal responses are overall stronger for more arousing sounds (compared to neutral), whereas startle reflexes, facial muscle activity, and heart rate, all responded more strongly to negative, relative to positive sounds.

In terms of linking emotions to other cognitive processes, Bradley and Lang (2000)

[1] We will return to this issue, and show some similar findings later in the current work.

also found that recall performance was also better for more arousing stimuli stimuli, similarly to electrodermal reactivity. Finally, based on measures of self-report, these authors concluded that the affective sounds in the IADS-2 database present a Valence $\times$ Arousal relationship which is similar to that of image stimuli from the IAPS.

Research involving affective sounds bears some formal similarity to studies relying on affective words, in that both have been used as research tools for similar purposes. As such, following the same approach as for ANEW, work was carried out to compare the IADS-2 database norms established on American samples, with other cultures. For example, Redondo, Fraga, Padrón, and Piñeiro (2008) present results comparing the IADS-2, ANEW and IAPS norms to data collected from Spanish samples, and identified several cultural differences: e.g., in terms of Valence, the American sample rated words and sounds more positively than the Spaniards, but no differences were found for images; however for Arousal, all the three stimulus types were found to be overall more arousing for the Spanish raters than for the American raters.

Another parallel can be drawn with the work carried out on the ANEW word database, given that Stevenson and James (2008) also attempted to classify the IADS sounds, and arrived at the same conclusion as before - that rating stimuli according to the PAD model, or conversely, classifying them into basic emotions, are approaches that are not reducible to one another, with each approach presenting a unique view of emotions. This is concluded based on the fact that the emotional dimensions did not consistently predict categorical data (or vice versa).

Finally, even the investigation of psychopathy has been attempted with sounds, just as it has with words, and again confirmed the differences between individuals with high psychopathy scores and normal controls (Verona, Patrick, Curtin, Bradley, & Lang, 2004).

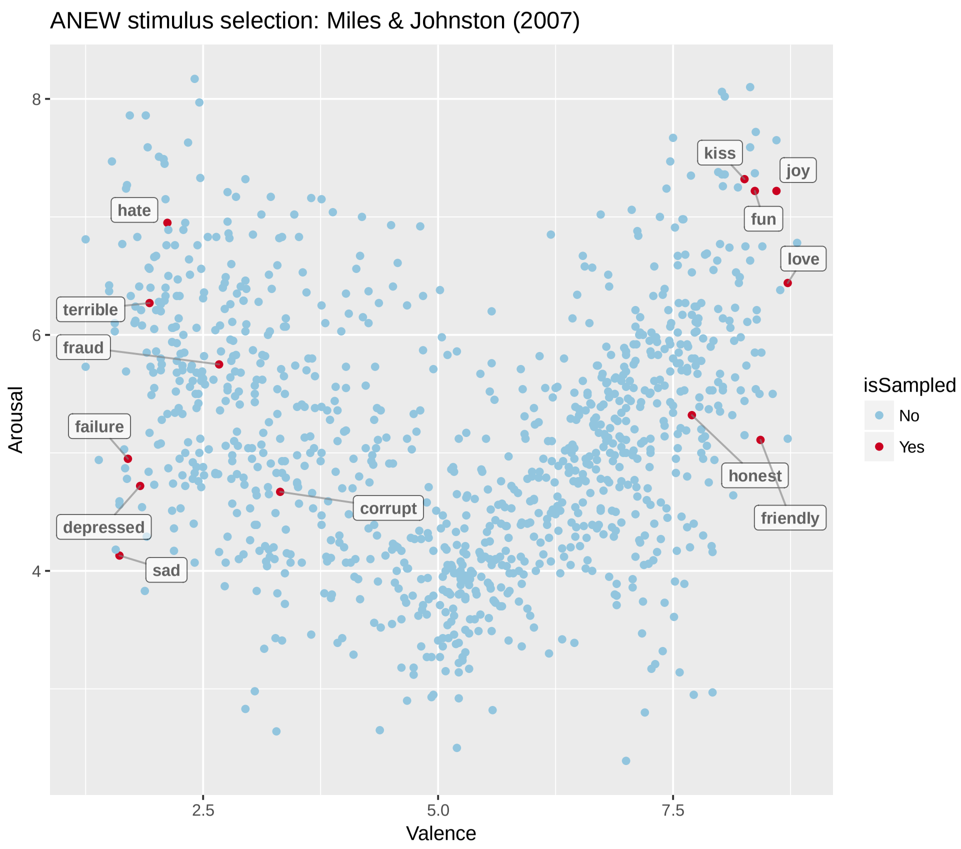
3.1.4 Common stimulus selection strategies

To explore differences in self-report between multiple emotion elicitation methods, and in order to minimise error, it is necessary to use a consistent method for sampling stimuli across *all the included modalities*, e.g., when using the IAPS, ANEW or IADS-2 databases.

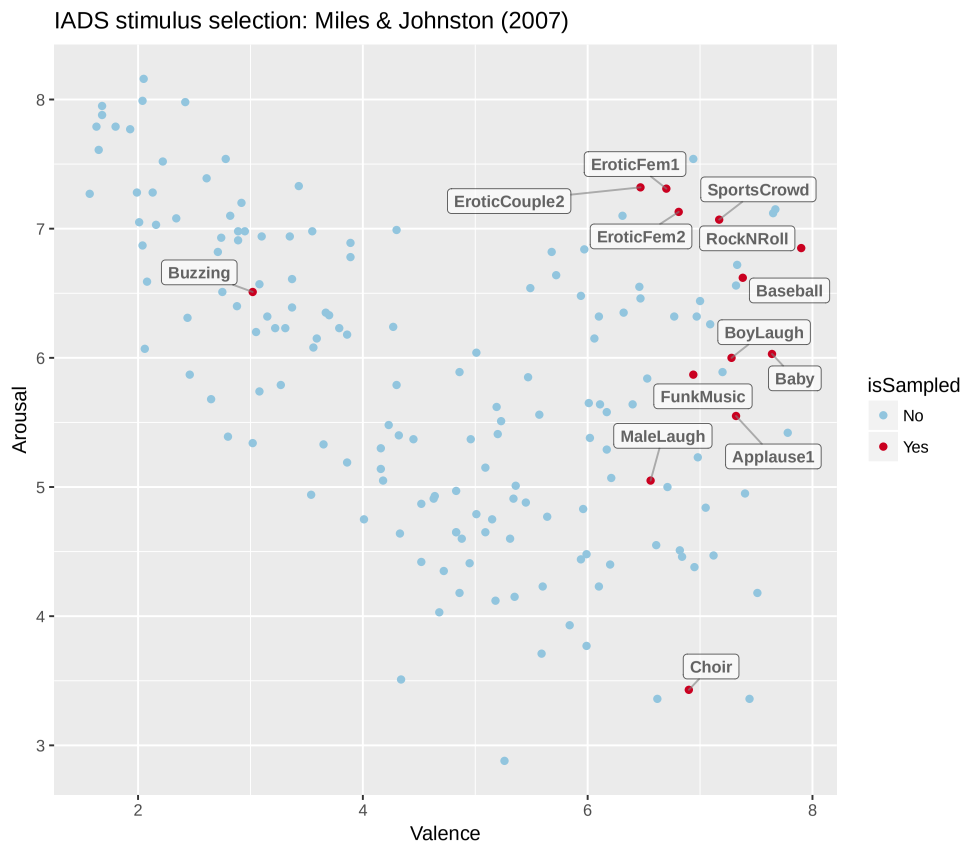
Despite this, even when using a single modality, studies often rely on very heterogeneous stimulus selection methods, which usually take into account only Valence and (at most) Arousal levels, but not the Dominance dimension as well. In addition, as shown in the previous chapter, such studies routinely adopt "manual" methods for sampling stimuli, which likely do not constitute optimal ways of partitioning them into separate levels of a treatment variable (i.e., "type of emotional content"). This would become particularly problematic when dealing with multiple stimulus modalities at the same time.

For example, Lewis, Critchley, Rotshtein, and Dolan (2007) used positive and negative words from the ANEW database to find a case of double dissociation between the dimension of Valence (related to the orbitofrontal cortex) and Arousal (related to activation in the amygdala). However, stimuli were sampled using a relatively unrefined method, i.e., positive words were considered to be those with Valence levels over a score of 7 on the 9-point Likert scale used, whereas negative words were considered to be those below a score of 3. Similarly, Scott, O'Donnell, Leuthold, and Sereno (2009) sampled positive words with Valence scores above 6, and negative words below a score of 4, with neutral stimuli situated in between these cut-offs. In the same vein, words used by Miles and Johnston (2007) were considered positive if over a score of 7.5, in contrast to scores under 2.5, which were considered to represent negative stimuli.

By this point, the reader may of course have noticed that these cut-off points: are somewhat arbitrarily chosen; reflect an assumption (which is not validated empirically) that there should be three groups of stimuli (not more, nor fewer); and they are only useful for selecting stimuli based on a single dimension (usually Valence). This type of approach is illustrated in Figure 3.1, on page 104, alongside relevant caveats concerning the structure of the stimulus norms, which was not fully taken into account when sampling stimuli.



(a) Miles and Johnston (2007) word sampling. Some words cited by the authors as being part of the ANEW database (Bradley & Lang, 1999a) could not be found, and therefore are not labelled above. Because stimulus selection in this study is based on Valence only, other affective dimensions may vary without being taken into account (e.g., Arousal in this case).



(b) Miles and Johnston (2007) sound sampling. ***Note:*** Confusingly, one positive sound sampled by the authors from the IADS, and declared to have the label "cardinal" and ID code "116", appears to be a negative sound, labelled as "buzzing" in the IADS-2 database (Bradley & Lang, 2007b).

Figure 3.1: Example of a stimulus sampling strategy for two modalities: words and sounds sampled from the ANEW and IADS databases. ***Note:*** The authors had access to the same ANEW version as Miles and Johnston (2007), but only the IADS-2, in contrast to Miles and Johnston (2007), who appear to have used the IADS, first version.

On the other hand, Estes and Adelman (2008) and Janschewitz (2008) used a different strategy discussed in our previous work (Constantinescu et al., 2016): discretisation and crossing the Valence and Arousal dimensions. This is also a problematic approach, which implicitly (and inappropriately) assumes that Valence and Arousal are independent dimensions. The same technique was also used by Noulhiane, Mella, Samson, Ragot, and Pouthas (2007) to sample affective sounds from the IADS, by crossing the Valence and Arousal dimensions, in order to investigate how the emotional properties of sound stimuli affect time perception (e.g., negative sounds were perceived as lasting longer than positive sounds). From Figure 3.1 (page 104), however, we have seen that Valence and Arousal are not independent dimensions within these stimulus databases - rather, they show a quadratic relationship. As such, trying to sample stimuli from them as if Valence and Arousal were orthogonal can have unknown consequences, and may bias study results.

Often enough, the stimulus selection strategy may go altogether unreported, although the stimulus ID codes and stimulus group means on the PAD dimensions will usually still be specified. For example, Keil et al. (2007) investigated how affective stimuli may affect the processing of concurrent stimuli, in the form of startle probes. The authors found that participants react less to the startle probes when the foreground stimuli are emotionally valenced (positive or negative), compared to neutral. However, in Figure 3.2, on page 106, we can easily see that in this study, the chosen neutral and positive stimulus groups overlap to a large extent, with precise consequences for interpretability remaining unknown. Finally, there are exceptions, however, e.g., Hamann & Mao, 2002, who do not provide information about the strategy of stimulus sampling, nor the stimulus codes used.

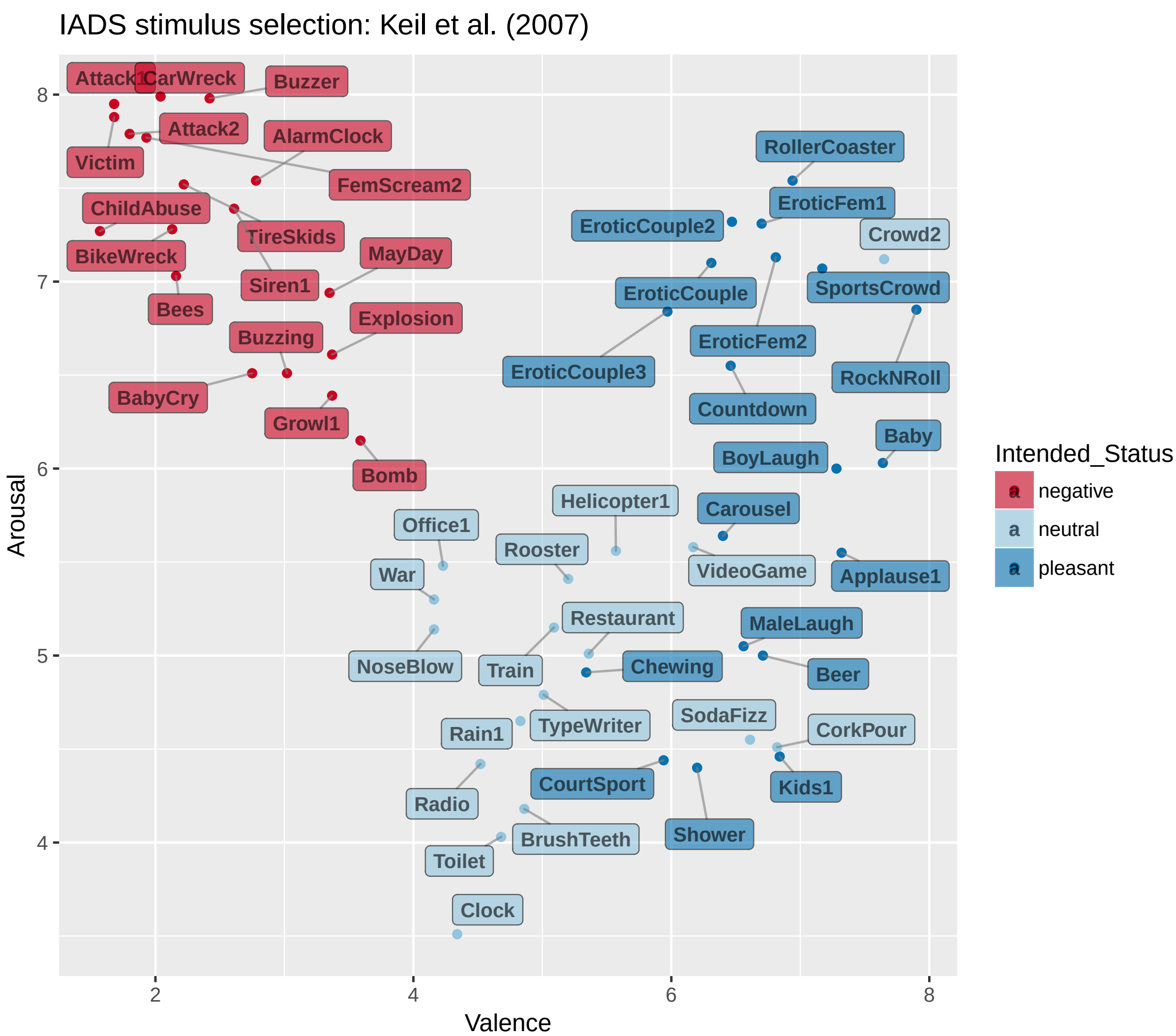


Figure 3.2: IADS sounds sampled by Keil et al. (2007) according to an unknown strategy. As can be seen, the group of neutral and positive/pleasant stimuli partially overlap. Strikingly, points such as "CorkPour" and "Kids1" were placed in different stimulus groups, although they occupy almost identical places in this 2D space.

3.1.5 Research aims

According to Bradley and Lang (2000) and Redondo et al. (2008), it is important to afford researchers the possibility to compare different stimulus modalities (i.e., images, words, sounds), in the hope that any similarities identified might provide information about a "common emotional system". This could be achieved by first sampling "images, words, and sounds with similar values" on the PAD model (Redondo et al., 2008, p. 787), and then comparing them in empirical research, in terms of self-report, behaviour and/or physiological measures.

In the previous chapter, and in line with this aim, we proposed a standard method for selecting affective image stimuli, based on model-based cluster analysis. Given that images are far from being the only emotion elicitation method used in controlled environments (with other popular options including affective words and sounds), and that few studies directly assess differences / similarities between multiple emotion-elicitation methods, the current work is intended to support further research into these topics.

Therefore, here we will be extending our stimulus sampling strategy from IAPS images, to two more stimulus modalities: words and sounds. This strategy will again use clustering algorithms to automatically identify the appropriate groups of stimuli, but this time, only once the different modalities have been matched together as closely as possible. As before, our method will take into account the structure of the databases and the full PAD model when forming the stimulus groups - which, as far as we are aware, has not previously been done in the literature.

3.2 Method

3.2.1 Data and instruments

In the current work, we resorted to three major stimulus databases, which include normative ratings for the three PAD dimensions: Valence (Pleasure), Arousal and Dominance, as well as standard deviations for the response distributions where the norms originated from. Approximately 100 American students rated each stimulus from the IAPS and IADS-2; however, for ANEW, the sample sizes used to produce the norms are unclear. We assume them to be similar to those from the other two databases, given they are produced by the same research lab, and that participants were tested in groups including between 8 and 25 students at a time (Bradley & Lang, 1999a), hence it is conceivable that multiple such groups could have been run for the same stimuli.

The number of stimuli in each database are as follows:

- The International Affective Picture System (**IAPS**): $N = 1182$ unique cases (Lang et al., 2008).

- The Affective Norms for English Words (**ANEW**): $N = 1030$. This older version of ANEW was used (Bradley & Lang, 1999a), given it also included word frequencies used later in this work as covariates. In the more recent version of the ANEW (Bradley & Lang, 2010), word frequencies are absent.

- The International Affective Digitized Sounds-2 (**IADS-2**): $N = 167$. The IADS-2 is thus the smallest database of the three (Bradley & Lang, 2007b).

The Affective Norms for English Text database (**ANET**, $N = 120$, Bradley & Lang, 2007a) was also consulted for explanatory purposes, and will be discussed later.

3.2.2 Analysis overview

Firstly, the data included within each database was cleaned in terms of: duplicate cases (if any were found), missing values (handled using listwise deletion), outlying values and cases where 95% confidence intervals around the norms spanned more than 1 point.

Secondly, the remaining data was combined across modalities, while adding an extra indicator column coding the stimulus modality: image, word or sound. Based on this mixed database, firstly we matched the sounds to the images available in terms of their PAD scores, and separately - the same sounds to the words. Afterwards, we observed which individual sounds were able to be matched to an image as well as a word, and retained these stimulus triplets for further use.

Thirdly, we applied various clustering algorithms to this mixed-modality dataset (composed of matched images, words and sounds) and, similarly to the previous chapter, model-based clustering was used to provide the final classification of the stimuli. Within each cross-modality / mixed cluster, a constant number of 5 images, words and sounds (those with the lowest clustering uncertainty) were sampled as the best representatives for their cluster.

Finally, cluster validation techniques were used to explore the clustering solution, and ANET texts were used additionally as a means to gain more insight into the meaning of the clusters discovered. A fifth "neutral" cluster was also formed separately as an artifice, for purely theoretical reasons - with details provided below.

3.3 Results

3.3.1 Data cleaning

Given that the three stimulus databases include hundreds of stimuli, a selection procedure was devised in order to select only a few more "desirable" cases from each modality. As such, we inspected each database, and first noted the existence of 12 duplicate cases within the IAPS, as discussed previously (see Table 2.1), with no such duplicates discovered within the IADS-2 or ANEW. We considered it necessary to remove these IAPS duplicates in order to avoid biasing subsequent analyses. Consequently, they were replaced with a single case defined by the averaged Valence, Arousal and Dominance scores.

After this step, the cases which presented any missing values were also removed, to later allow for a complete case analysis. This led to a reduction in sample sizes of 240, 58, and 0 cases for IAPS, ANEW and IADS-2, respectively (i.e., leading to: $N_{IAPS} = 942$; $N_{ANEW} = 972$; $N_{IADS-2} = 167$).

For ethical reasons, this was followed by the identification and removal of univariate outliers that spanned more than 2.5 median absolute deviations (MADs) around the median - a robust method of identifying outliers (Leys et al., 2013). The highest number of stimuli to be excluded using this criterion belonged to the IAPS database: 32 images, which had excessive values on the Dominance dimension (leading to an updated $N_{IAPS} = 910$). As far as ANEW is concerned, 3 and 4 words were removed due to their outlying scores on the Arousal and Dominance scales, respectively (updated $N_{ANEW} = 965$). Finally, no outliers were found in the IADS database using this method.

Additionally, one further measure was implemented to ensure the reliability of the stimuli to be used: building 95% confidence intervals around the stimulus norms. The selection of this measure was restrained by the type of data provided with the databases: just means and standard deviations. Thus, we aimed to select stimuli with confidence intervals spanning no more than 1 point around the stimulus norms, which we considered to be a sufficient degree of precision, considering that the 3 dimensions were measured on 9-point Likert scales. Using this criterion, cases judged too imprecise were removed (i.e., 61 for IAPS, updated $N_{IAPS} = 849$; 318 for ANEW, updated $N_{ANEW} = 647$; and 3 for IADS-2, updated $N_{IADS-2} = 164$), as this could affect inferences subsequently in our study. The final distributions are portrayed in Figures 3.3 (page 110) and 3.4 (page 111).

3.3.2 Matching the stimuli

Using `R` package `optmatch`, version `0.9-7` (Hansen & Klopfer, 2006), the IADS-2 sound database ($N = 164$ cases) was used as a pivot for the matching process. This is because, in order to create mixed-modality trios containing stimuli matched in a ratio of `1:1:1`, the number of such trios is necessarily constrained by the smallest database available, i.e., IADS-2. Thus, sounds were first matched to images ($N = 849$ cases), and separately, also to words ($N = 647$ cases). The matching distances used were based on all the three available PAD dimensions (using Mahalanobis Distance).

The `optmatch` algorithm creates a treatment by control-case matrix, where each cell represents the multivariate distance between two data points. A control is found for

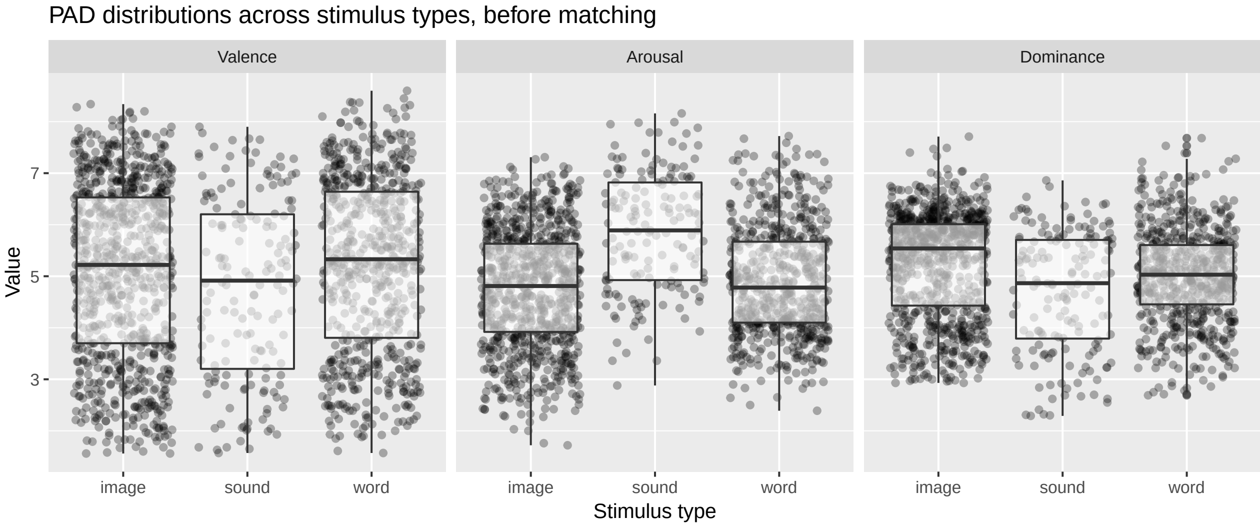


Figure 3.3: The boxplots above give some insight on the PAD distributions within the three databases. Data points have been jittered in the background to explain the shape of the boxplots, and reveal the amount of data available in each case. For example, the median for image Dominance scores is lifted by the heavier cloud of points which also suggests the presence of negative skew. Overall, the distributions show some differences exist between the modalities, which will be addressed in the stimulus matching process below.

each treatment-case (e.g., a matching image for every sound), based on which pair in the matrix shows the smallest Mahalanobis distance (Fredrickson, 2010). Other options for matching (e.g. stratified matching) were not considered, given that the only common measures between the stimulus databases were just the PAD dimensions, and not any additional variables.

In the case of sound-image matching, it was possible to use *pairwise* matching and the `pairmatch()` routine, given the large amount of the IAPS stimuli still available after data cleaning, relative to IADS sounds. This means that each of the 164 sounds found exactly 1 image to be paired with, with the remaining excess images being removed. Illustrative examples include image `1110` ("`Snake`") being matched to sound `106` ("`Growl 1`") based on PAD values, and image `4559` ("`Romance`"), matched to sound `351` ("`Applause`"). Thus, even if the stimuli are semantically and content-wise quite different, they are extremely similar in terms of the type of emotional state they convey - which can prove to be a very useful quality to control in research.

In the case of sound-word matching, *pairwise* matching failed with an error[2], and `optmatch` help files revealed more information on the nature of this problem: "[...] matching can still fail, if there is too much competition for certain controls; if you find yourself in that situation you should consider full matching, which necessarily finds a

[2] `Error in pairmatch.matrix(m, controls = controls, data = mfd, remove.unmatchables = remove.unmatchables,: not enough controls in some subclasses.`

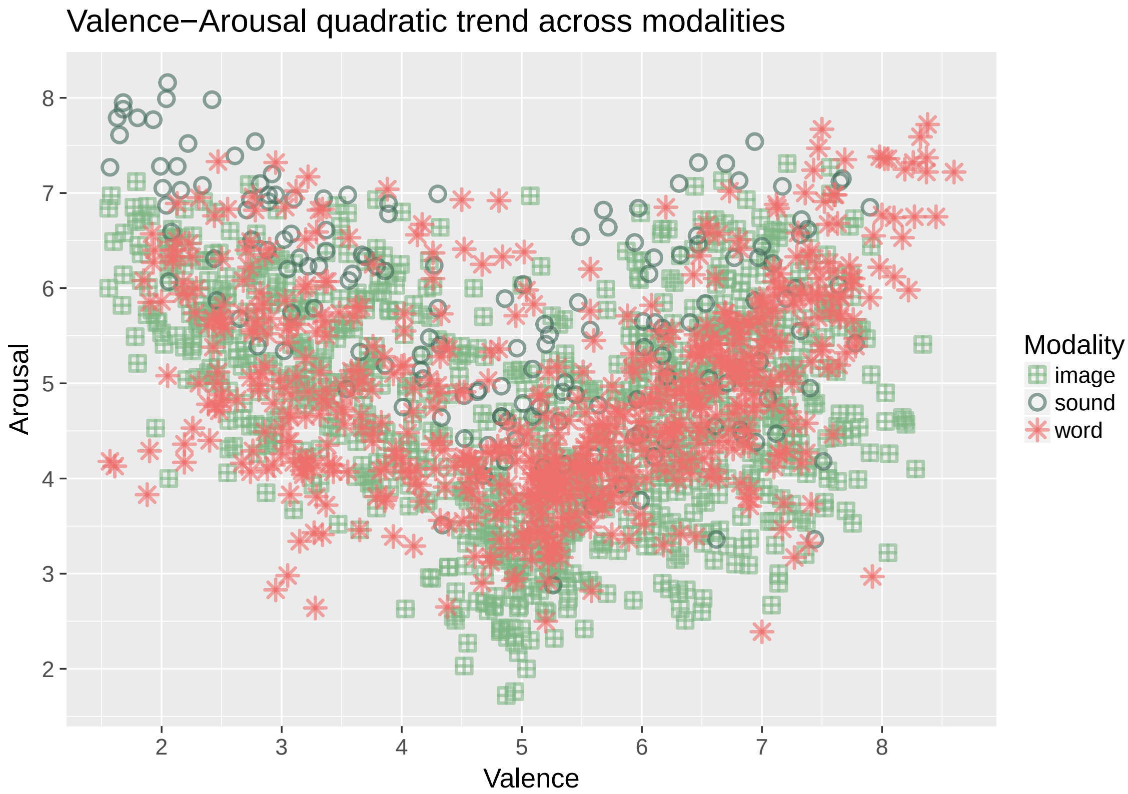


Figure 3.4: Quadratic trends across modalities: interestingly, all three modalities follow the same U-shaped distribution, when plotted in a 2D space defined by Valence and Arousal. It is also worth noting that sounds, in particular, occupy the higher levels of Arousal, relative to the two other modalities - which will introduce a certain amount of bias in the matching process.

match for everyone with an eligible match somewhere". Thus, *full matching* is not restricted to `1:1` matching, but can identify matches in various ratios of treatment to control-cases (e.g., 1 sound with 3 words, etc.), based on the distance measure used. Consequently, we indeed resorted to *full* matching, and the `fullmatch()` routine, for sounds and words, without specifying any constraints in the algorithm (such as the maximum number of controls allowed per treatment case, etc.), in order to find the best data-driven solution. The results led to a stratum structure which is displayed in Figure 3.5, on page 112.

Among the 164 sounds, only 79 could be paired up with a unique word (in `1:1` matches), with the remaining matches including either multiple sounds matched to 1 word, or multiple words matched to the same sound. In these cases, the Mahalanobis distance matrix was used to rank the multiple matching candidates from best (i.e., smallest distance), to worst (i.e., largest distance). Thus, based on these ranks, we were eventually able to identify `1:1` matches even from the `multiple:1` and `1:multiple` matches, and thus increase the number of usable stimulus pairs for further analysis.

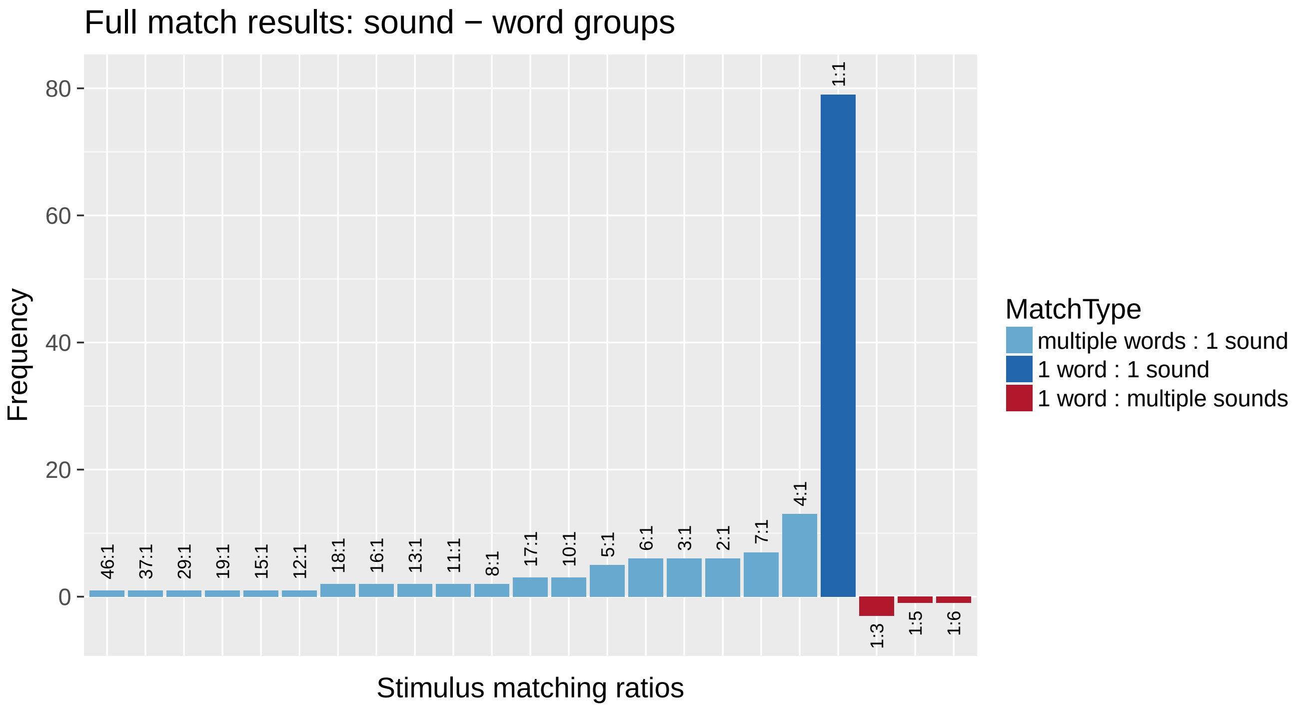


Figure 3.5: As shown above, most sounds found a unique match to one other word. However, due to high competition in the data, some sounds were also matched to multiple words, or conversely, the same word was matched to multiple sounds.

Examples from our results include pairs such as: sound `295` ("`Couple sobbing`") being matched to word `596` ("`knife`"), or sound `201` ("`Erotic Fem 1`") being matched to word `152` ("`excitement`").

Overall, this process led to a number of 164 sound-image pairs, and 149 sound-word pairs. Despite starting with an equal number of to-be-matched sounds, it is worth noting that, specifically for the cases where multiple *sounds* were matched to *the same word*, the overall number of stimulus pairs had to be reduced across all the modalities. This is because sounds could not be allowed to re-occur within our matched datasets. Thus, a total of 149 trios (of sounds, with both their matching image *and* word counterparts) were left for further analysis (i.e., $N = 447$; also see Appendix B on page 441 for a list of these stimulus trios).

We consider the matching procedure to have been a success, given the clear improvement in similarity across distributions, shown in Figure 3.6 (page 113), relative to the previous Figure 3.3 (page 110).

Before the clustering process which will follow, the trio data is further displayed in Figure 3.7, as a parallel coordinate plot. This breaks down the three-dimensional data structure, by connecting the coordinates of each stimulus, across the multiple axes. Interestingly, this shows that some of the stimuli have opposing trajectories in this space,

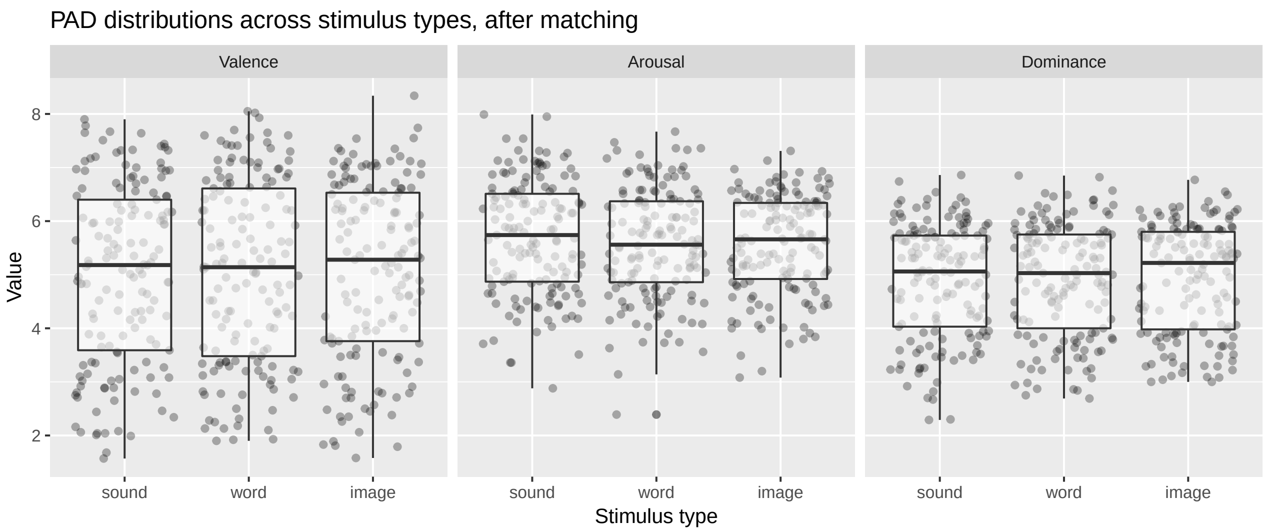


Figure 3.6: Above, visibly less data contributes to the boxplots than before, but instead, the medians and quartiles are located around very similar levels, leading to a fairly striking difference when comparing to the pre-matching distributions.

e.g., some stimuli low in Valence are high in Arousal, and vice versa. Ideally, the cluster analysis will detect these features, and use them to place such stimuli in different groups - which will be investigated in the following section.

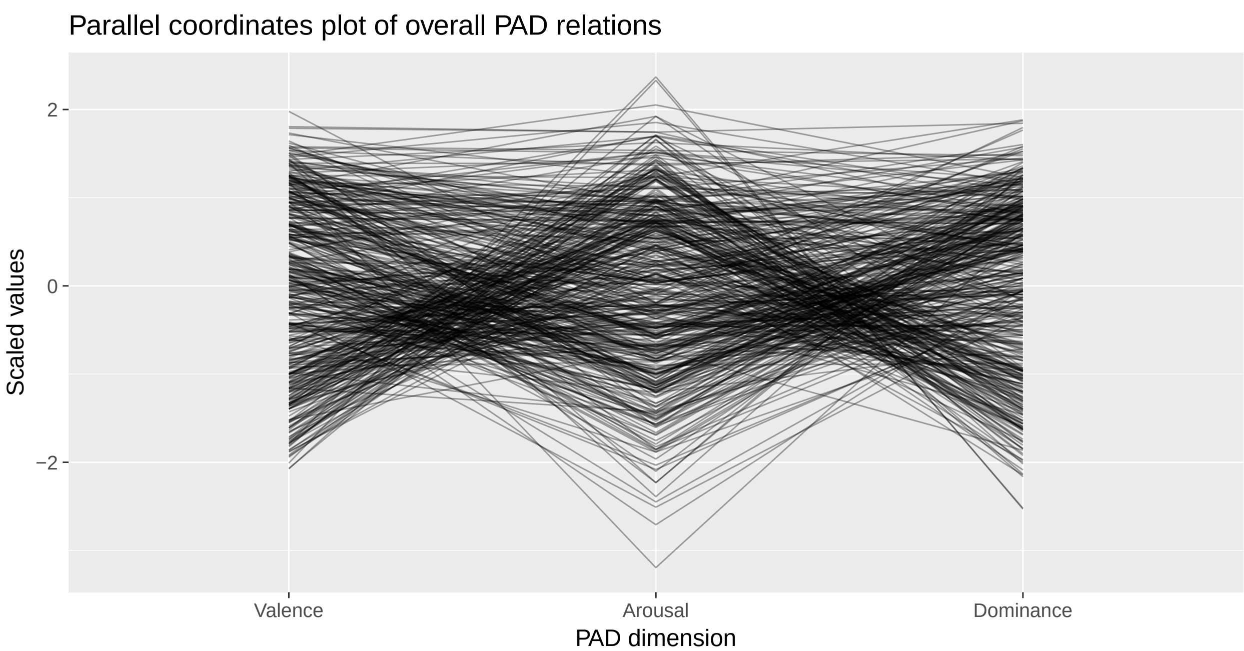


Figure 3.7: Various trends in the mixed-modality data (sounds, images, and words matched together), shown across the PAD dimensions. Each line connects the coordinates of a single data point.

3.3.3 Classifying the data

Given that the matched dataset was balanced in terms of the stimulus modalities included (i.e., with every modality typified by exactly 149 items), we were able to run a cluster analysis without fear of any modality being over-/under-represented in the results, and therefore risk biasing them. Our aim is to reveal any structure these data might present, and based on this, create stimulus categories serving as the levels of an independent variable in future research.

Before proceeding, it is worth mentioning that, in the correlation matrix of the PAD dimensions, Valence and Dominance are highly related (see Table 3.1). Given our previous arguments about preserving the "natural" relationships between variables when selecting database stimuli (Constantinescu et al., 2016), this is not surprising, and should be taken into account when creating stimulus groups.

Table 3.1: Correlation matrix for Valence, Arousal, Dominance within the matched dataset.

	Valence	Arousal	Dominance
Valence	-	[-0.304, -0.128]	[0.901, 0.931]
Arousal	-0.218^{***}	-	[-0.401, -0.234]
Dominance	0.917^{***}	-0.320^{***}	-

$^{***}p < 0.001$; $^{**}p < 0.01$; $^{*}p < 0.05$. The values above the diagonal represent 95% confidence intervals for the r coefficients below the diagonal.

However, according to authors such as, e.g., Mooi and Sarstedt (2011, p. 242), such characteristics in the data are known to affect clustering solutions by giving more weight to the correlated variables than the rest, when forming a solution. The same source recommends to either (re)consider the theoretical relevance of the clustering dimensions, or to first compute a factor analysis, and use just the factor scores in the subsequent cluster analysis, instead of the several correlated variables.

In our case, we considered that there are valid reasons for entering Valence, Arousal and Dominance into the cluster analysis, as other research has shown they can independently explain portions of variance in emotional responses (Bradley & Lang, 1994). This suggests that Valence and Dominance may be strongly correlated just for these particular stimuli, rather than necessarily in general. As such, it was judged more appropriate to retain the correlated dimensions for analysis, rather than: perhaps exclude Dominance, or alternatively, retain just Arousal and a composite measure of both Valence and Dominance.

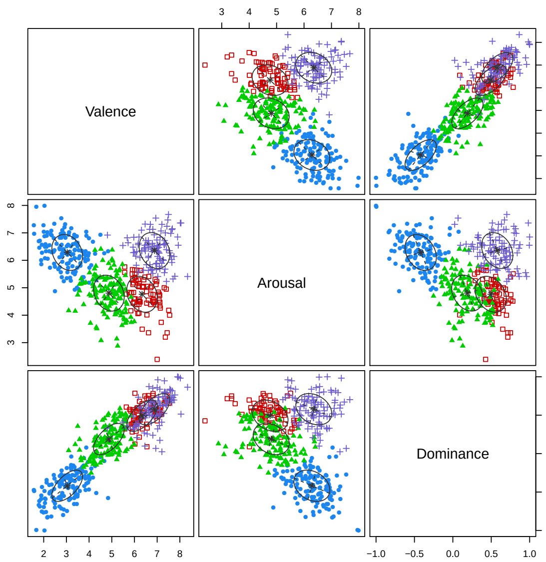
Various clustering algorithms were tested with these dimensions, notably including model-based clustering, as previously, but also hierarchical and k-means clustering, with results reported in the next subsection.

3.3.3.1 Model-based clustering

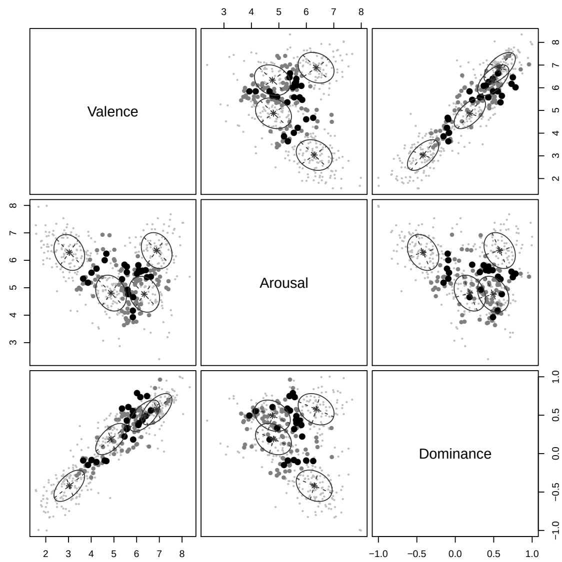
This solution was computed using the `R` package `mclust`, version `4.3` (Fraley & Raftery, 2006), and indicated that the best fitting model was reached for 4 ellipsoidal clusters, with equal volume, shape and orientation (model coded as 'EEE'), with $BIC = -3249.997$; Log-likelihood $= -1560.922$ (and $N = 447; df = 21$). For this BIC-optimal model, cluster centroids and mixing proportions are specified in Table 3.2, and two graphs detailing the classification and level of uncertainty are presented as part of Figure 3.8.

Table 3.2: Cluster centroids and mixing proportions for the model-based clustering solution. Mixing proportions express the cluster N as a proportion out of the total N.

Cluster	Valence	Arousal	Dominance	N	Mixing proportion
1	3.04	6.29	3.62	136	0.31
2	6.33	4.76	5.70	77	0.19
3	4.87	4.80	5.01	132	0.28
4	6.88	6.36	5.89	102	0.23



(a) Stimulus classification into the 4 clusters.



(b) Stimulus membership uncertainty: the larger and darker the points, the higher the uncertainty.

Figure 3.8: Mixed-modality clusters created using model-based clustering.

The second best-fitting model was also 'EEE', but with 3 components (and a considerably smaller BIC value: $BIC = -3254.176$), whereas the worst fit was achieved for

1 component, for either the 'EII' or 'VII' model (i.e., spherical, and equal vs. unequal volume): smallest $BIC = -4477.502$[3].

The optimal clustering solution was validated using several methods, all of which suggest that this 4-cluster solution is very robust.

Firstly, we created two random halves of the matched dataset, and computed independent clustering solutions for each of them, both set to extract 4 clusters, reflecting the solution achieved on the full dataset. Subsequently, each clustering solution (from its respective halved dataset) was used to predict the solution independently achieved on the other half, and then the degree of association was computed between the prediction and the actual clustering solution, which had previously been calculated. This was done under the assumption that if the $k = 4$ solution is well-supported, then these associations should be strong. This was indeed found to be the case, with an average computed Cramer's φ of 0.932 (i.e., an extremely high degree of association between the actual classifications and predictions).

Secondly, we computed two measures of overlap between these actual and predicted classifications: the Rand and Adjusted Rand Index. Both indicated very high levels of similarity, i.e., an average Rand index of 0.955 (where 1 represents perfect overlap), and an average Adjusted Rand Index (i.e., a version of the Rand Index which has been penalised for any randomness in the classification) of 0.882, which is still sizeable.

Thirdly, the same halved datasets were used for computing the normalised version of the Variation of Information index (VI, Meilă, 2007), i.e., the VI values divided by the log of the sample size used. This is a measure for how much information is lost / gained when moving from the predicted to the actual classifications. On average, these normalised indices equalled 0.065, i.e., very little information is either gained / lost when moving from one classification to another. In other words, they are very similar.

In addition, we also iteratively removed a random 10% of the data points, and reassessed the BIC-optimal number of clusters each time, over 1000 repetitions. Across all the repetitions, 73.6% of the time, the same 4-cluster classification emerged, suggesting this is a stable solution. Other far less frequent options included 2 clusters (0.8% of cases), and 3 clusters (25.6%).

The last method employed to verify the effectiveness of our stimulus selection procedure was the intra-class correlation coefficient (ICC), here referring to the degree of similarity for cases *within* the same cluster, rather than inter-rater reliability - the other major application for this measure (Field, Miles, & Field, 2012, p. 859). ICCs represent the ratio between the variance between groups, and the total variance in the data (i.e., between *and* within groups), and therefore can vary between 0 and 1. The ICCs were

[3] As before, BIC values in the `mclust` package are reversed, so that the ideal BIC value is the largest, rather than the smallest.

computed using the R package ICC, version 2.3, with CIs of the type "THD", which are suitable for unbalanced data (Wolak, Fairbairn, & Paulsen, 2012). In Table 3.3, we can see that for each PAD dimension considered separately, but also the interaction term between them, the values of the estimated ICCs are fairly high.

Table 3.3: Intra-class correlations for the four clusters including images, words, and sounds.

	ICC	95% CI	Groups	Within variance	Between variance
Valence	0.90	[0.73; 0.99]	4	0.38	3.25
Arousal	0.68	[0.41; 0.97]	4	0.39	0.84
Dominance	0.86	[0.66; 0.99]	4	0.19	1.19
PAD interaction	0.90	[0.73; 0.99]	4	827.24	7087.88

In the case of the PAD interaction, for instance, these results suggest that our selection procedure was extremely efficient, with around 90% (or at least 73%) of the total variability seen in our data being due to between-group differences, whereas the remaining 10% (or 27%, in the worst case) being attributable to within-cluster variability. So regardless of the fact that there are different modalities included in each cluster, the very high ICCs also suggest our matching procedure was efficient.

Finally, the assumption of multivariate normality was checked for each cluster, both visually and using a series of tests, e.g., kurtosis values appeared more problematic than skewness, according to the Mardia test. However, this was not seen as a major cause for concern, given two reasons.

Firstly, we used bootstrapping to alleviate concerns with non-normality of the data. On generating 1000 bootstrapped samples[4], and using these to build 95% confidence intervals (CIs) around the previously computed cluster centroids, we noted that the CIs were quite narrow around the centroids (see Figure 3.9, p. 118). This suggests that non-normality was not influencing cluster estimation to any important degree.

Secondly, eigenvalues representing the amount of variation in space associated with each component / dimension in 3D space, tended to reflect ellipsoidal shapes (as earlier suggested by the 'EEE' model) - which is a desirable feature in terms of assuming multivariate normality. These results suggest that the data can be considered normal for the purposes of this analysis.

3.3.3.2 Other clustering methods

Despite the robustness of the model-based clustering solution, we were also interested in checking further whether two other algorithms converge on a similar number of clusters:

[4] In order to do this, it was necessary to upgrade the mclust package to a newer version (v.5.2.3), which included a bootstrapping option.

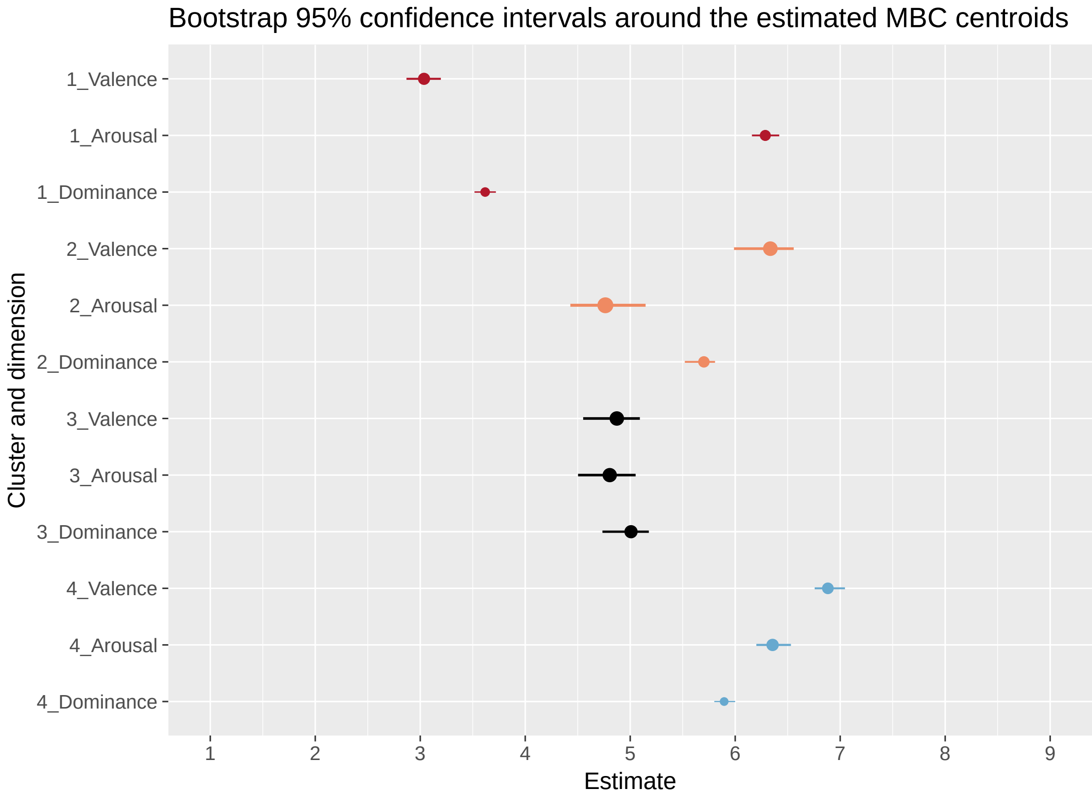


Figure 3.9: 95% bootstrapped confidence intervals for cluster centroids estimated using model-based clustering (MBC). Point size is proportional to standard errors of the estimates (i.e., centroids).

hierarchical clustering and k-means.

In the case of hierarchical clustering, we used cophenetic correlations, Gower distances and Shepard-like diagrams to discover the most suitable combination of distance and linkage measure (for details on all these criteria, please see the paper and supplementary material by Constantinescu et al., 2016), as well as Chapter 2. Based on these measures, it emerged that correlation (rather than Euclidean) distances, with either Ward or average linkage reproduced the structure of this dataset most closely.

Using these combined distance and linkage methods, we computed further criteria designed for use with hierarchical clustering, to get an indication of the most appropriate number of clusters in the data. These are detailed in Table 3.4.

When using k-means, various indicators were also checked for identifying what number of clusters (k) is present within the data. These included the Calinski-Harabasz Index and the Simple Structure Index (SSI), as well as others - all detailed in the same Table 3.4, on page 119.

Because of the general lack of consistency which is apparent in this table, it was decided to retain the far more robust and stable results derived from model-based clustering.

Table 3.4: Various hierarchical clustering criteria, and their recommended value for k.

Method of selecting k		`R PackageName::function()`	k
Clustering type : Hierarchical			
Ward-type plot[a]		Custom function	4
The "Elbow" method		`GMD::elbow()`	5
Internal validation and stability		`clValid::clValid()`	2 & 8[d]
Average Silhouette Width[b]		Custom function	6
Graphs of fusion level values[b]		Custom function	5
Mantel Optimality[b]		Custom function	3
Clustering type : K**-means**			
SSI criterion		`vegan::cascadeKM()`	8[d]
Calinski criterion		`vegan::cascadeKM()`	2
Internal validation and stability		`clValid::clValid()`	2, 6, or 8[d]
Random halves SSI	half 1	`vegan::cascadeKM()`	8[d]
	half 2		8[d]
Random halves Calinski	half 1	`vegan::cascadeKM()`	3
	half 2		2
χ^2 of halves (correlation distances)[c]		Custom procedure	8[d]
χ^2 of halves (Euclidean distances)[c]		Custom procedure	8[d]

[a] Based on `R` code from: `http://www.statmethods.net/advstats/cluster.html`.

[b] See Borcard et al. (2011, p. 53 - 114) and the associated `R` code.

[c] The relevant `R` code is available from the author's GitHub repository: `https://github.com/CaterinaC/ANEW_IAPS_IADS_match_and_cluster`.

[d] Also the largest number of clusters tested.

3.3.4 Selecting best representatives

In a similar fashion to our previous work (see Chapter 2, or Constantinescu et al., 2016), and given that model-based clustering also provides uncertainty levels for the membership of each stimulus to its cluster, we were able to rank the stimuli in each cluster according to the likelihood of "correct" classification. We thus sampled the first 30 most likely cases per cluster (although this number could be increased, depending on researchers' aims), and examined the modalities within these subsets in Table 3.5:

Table 3.5: Modality counts within the 30 best representatives from each cluster.

	Cluster 1	Cluster 2	Cluster 3	Cluster 4
sound	13	10	11	13
word	9	9	9	12
image	8	11	10	5

Given that the minimum count in the table occurs for images within the fourth cluster, we decided to sample the same number of stimuli across each modality and cluster, i.e., the 5 images, words and sounds from each cluster which present the lowest level of uncertainty. These are listed individually in Table 3.6:

Table 3.6: The 5 most representative images, words and sounds from each cluster.

No	Valence	Arousal	Dominance	Type	Code & description	Uncertainty	Cluster
1	1.680	7.950	2.300	sound	279_attack1	0.000	1
2	2.040	7.990	2.290	sound	424_carwreck	0.000	1
3	1.990	7.280	2.820	sound	292_malescream	0.000	1
4	2.160	7.030	2.670	sound	115_bees	0.000	1
5	2.340	7.080	2.700	sound	420_carhorns	0.000	1
6	1.920	6.570	2.860	word	591_drown	0.000	1
7	2.470	7.330	3.220	word	15_ambulance	0.000	1
8	2.100	6.330	2.840	word	616_trauma	0.000	1
9	2.180	6.060	2.690	word	618_victim	0.000	1
10	2.780	6.820	2.940	word	604_scared	0.000	1
11	1.830	6.720	3.000	image	9412_deadman	0.000	1
12	1.790	6.700	3.040	image	3110_burnvictim	0.000	1
13	1.580	6.970	3.460	image	3130_mutilation	0.000	1
14	1.810	6.450	3.170	image	9187_injureddog	0.000	1
15	2.060	6.490	3.110	image	9414_execution	0.000	1
16	7.440	3.360	6.290	sound	809_harp	0.008	2
17	7.510	4.180	6.070	sound	810_beethoven	0.023	2
18	6.950	4.380	5.910	sound	150_seagull	0.036	2
19	7.120	4.470	5.730	sound	151_robin	0.037	2
20	6.840	4.460	6.070	sound	112_kids1	0.045	2

21	7.180	3.740	5.740	word	320_politeness	0.021	2
22	7.140	4.260	5.930	word	466_useful	0.025	2
23	6.890	4.500	5.700	word	796_humane	0.045	2
24	6.710	4.390	6.020	word	761_garden	0.051	2
25	7.090	4.770	5.790	word	786_heal	0.091	2
26	7.360	3.200	6.210	image	5200_flowers	0.012	2
27	7.550	4.000	6.170	image	2314_binoculars	0.014	2
28	7.120	4.340	5.820	image	2156_family	0.029	2
29	7.080	4.550	5.880	image	7480_pasta	0.049	2
30	6.840	4.560	6.020	image	7390_icecream	0.053	2
31	4.340	3.510	4.640	sound	708_clock	0.005	3
32	4.330	4.640	4.950	sound	382_shovel	0.015	3
33	4.520	4.420	4.930	sound	723_radio	0.016	3
34	4.520	4.870	5.040	sound	358_writing	0.025	3
35	4.680	4.030	5.620	sound	700_toilet	0.030	3
36	4.320	3.560	4.610	word	405_solemn	0.005	3
37	3.500	4.740	5.390	word	913_obnoxious	0.009	3
38	4.310	4.900	5.500	word	834_kick	0.018	3
39	3.390	4.150	4.850	word	806_immature	0.018	3
40	4.270	4.750	4.860	word	608_skull	0.019	3
41	4.530	3.080	4.550	image	9210_rain	0.004	3
42	4.060	3.710	4.460	image	2206_fingerprint	0.009	3
43	4.350	4.440	5.090	image	1230_spider	0.012	3
44	4.620	4.130	5.000	image	2410_boy	0.016	3
45	4.220	4.920	4.950	image	1240_spider	0.022	3
46	7.670	7.150	6.440	sound	817_bongos	0.000	4
47	7.900	6.850	6.860	sound	815_rocknroll	0.000	4
48	7.650	7.120	6.090	sound	311_crowd2	0.000	4
49	6.700	7.310	5.930	sound	201_eroticfem1	0.000	4
50	6.940	7.540	4.730	sound	360_rollercoaster	0.000	4
51	8.020	7.360	6.820	word	530_sexy	0.000	4
52	7.500	7.670	6.180	word	152_excitement	0.000	4
53	8.050	7.360	5.750	word	384_sex	0.000	4
54	7.470	7.470	6.110	word	422_surprised	0.000	4
55	7.430	7.240	6.390	word	512_erotic	0.000	4
56	7.210	7.310	4.630	image	8492_rollercoaster	0.001	4
57	6.670	7.130	5.730	image	4668_eroticcouple	0.001	4
58	6.870	6.930	5.670	image	4659_eroticcouple	0.002	4
59	6.990	6.740	5.850	image	4670_eroticcouple	0.003	4

After clustering, these trio data are also illustrated in Figure 3.10 (page 122), where the most representative stimuli from each modality are also labelled (with added jitter to avoid overlap). Compared to the previous Figure 3.7 (page 113), it is indeed obvious that the model-based clustering algorithm detected the different pathways illustrated, and placed the relevant cases into different clusters.

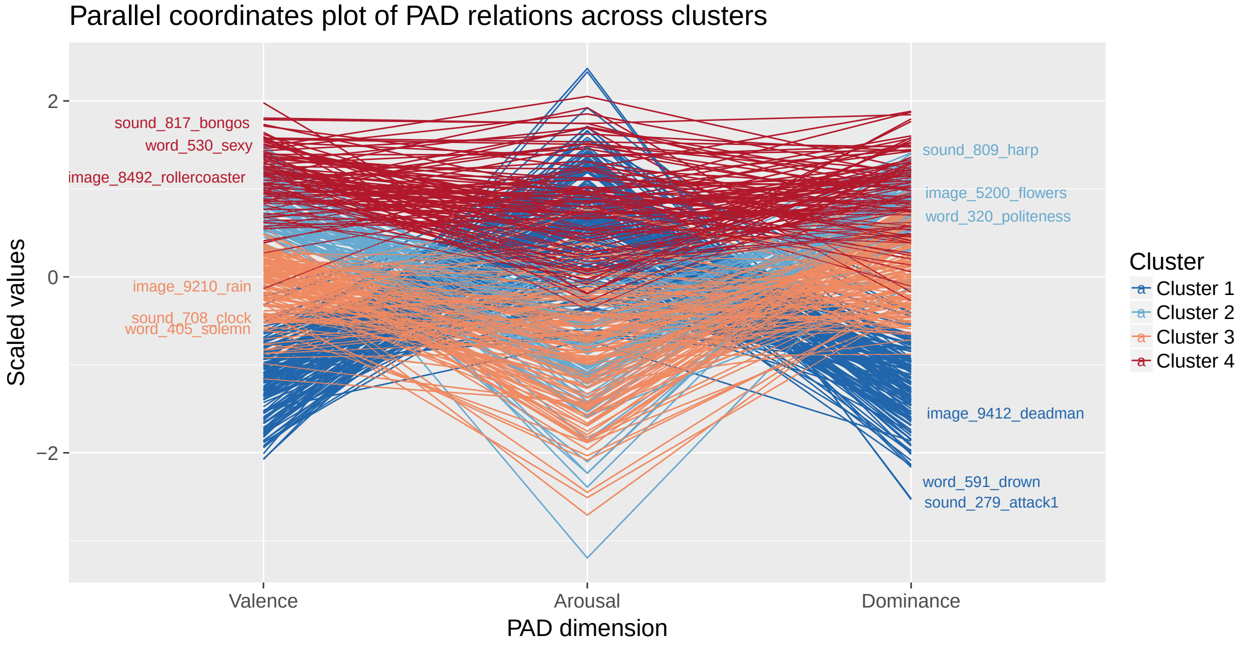


Figure 3.10: The four clusters show fairly clear, and different, relationships between PAD dimensions. The best representative from each modality, within each cluster, is printed on the sides of the plot.

3.3.5 Adding a fifth neural cluster

It is worth noting that the set of four symmetrical clusters (mildly negative; very negative and intense; positive and serene; positive and exciting) fails to include a neutral cluster which could be used as a baseline[5]. Even if it may be argued that emotional stimuli are very rarely "neutral", or that "neutrality" may not mean what one might assume in emotional stimuli, for theoretical purposes we "manually" crafted an additional cluster as a methodological artifice. When collecting empirical data using the stimuli nested in the four data-driven clusters, this additional cluster could serve as a term for comparison / baseline.

Any stimuli not included in the groups of best representatives for the four previous clusters, presented earlier in Table 3.6 (page 120), were considered as candidates for forming this additional cluster. The new cluster was constructed so that it would contain the five images, words and sounds which were closest to a value of 5 on all the three dimensions of the PAD model, since this is the midpoint on the 9-point Likert scales

[5] However, see our previous work in Constantinescu et al. (2016), as well as Section 9.2.2, p. 383, where we detail that "neutrality" in these stimulus databases is a vague concept, without a clear empirical definition.

used in the international databases. The resulting cluster can be viewed in Table 3.7, as well as Figures 3.11 and 3.12, on pages 124 and 125, where it is presented alongside the other clusters.

Table 3.7: Creating a neutral cluster based on Euclidean distances from a PAD vector of `(5, 5, 5)`, representing the centre of PAD 3D space. The stimuli from this cluster complete the set from Table 3.6 (page 120).

No	Valence	Arousal	Dominance	Type	Code & description	Euclidean distance from middle of 3D space
1	4.640	4.930	5.000	sound	130_pig	0.367
2	5.090	5.150	4.670	sound	425_train	0.373
3	4.960	5.370	5.060	sound	104_panting	0.377
4	4.830	4.970	4.660	sound	722_walking	0.381
5	4.830	4.650	5.070	sound	246_heartbeat	0.395
6	4.720	5.030	5.030	word	23_army	0.283
7	5.140	4.860	5.290	word	1004_swamp	0.351
8	4.810	5.360	4.910	word	957_razor	0.417
9	4.740	5.330	4.810	word	966_rough	0.461
10	5.300	4.620	4.880	word	434_theory	0.499
11	4.950	5.090	4.890	image	9422_battleship	0.151
12	4.990	5.140	4.740	image	1645_wolf	0.295
13	5.030	4.930	5.320	image	2220_maleface	0.329
14	5.280	4.880	4.750	image	1908_jellyfish	0.394
15	4.920	5.130	5.380	image	3550.2_coach	0.410

3.3.6 Cluster validation for the five-cluster solution

The fifth cluster was found to still differ significantly from the closest data-driven cluster (i.e., the mildly negative cluster) on Valence, Arousal, and the interaction of all three dimensions (although not on the Dominance dimension, on its own). Results are reported in Table 3.8, and confirmed visually by e.g., Figure 3.12 (p. 125), where these two clusters are shown to differ on the Valence and Arousal dimensions, but are indeed situated around the same level on the Dominance dimension.

Table 3.8: *t*-tests comparing two neighbouring clusters: a mildly negative, data-driven cluster, and the artificially constructed neutral cluster.

Dependent	Test	Results
Valence	Welch Two Sample *t*-test:	$t(21.28) = 6.33,\ p < .001$
Arousal	Welch Two Sample *t*-test:	$t(18.30) = 4.79,\ p < .001$
Dominance	Welch Two Sample *t*-test:	$t(24.71) = 0.01,\ p = .990$
PAD interaction	Welch Two Sample *t*-test:	$t(18.95) = 6.70,\ p < .001$

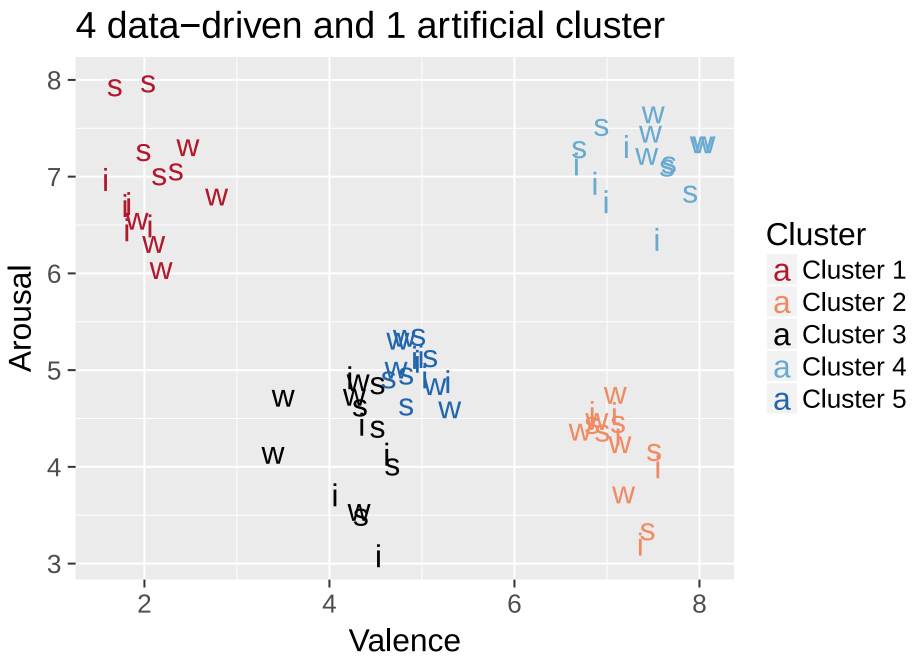


Figure 3.11: 2D plot of all the five clusters: the initial four resulting from the cluster analysis, and the fifth, built "manually" on theoretical grounds to act as a baseline / neutral cluster. Plotted letters represent the three modalities included: sounds, words and images.

The fifth cluster was also used as a reference group in a regression model, and found to significantly differ from the other four, on a composite measure describing the interaction of Valence × Arousal × Dominance. Moreover, we also inserted a modality factor into the model, to take into account any differences between words, sounds and images. As expected due to the matching procedure, there were no significant differences associated with this modality factor. The R^2 value is also very high, with $R^2 = 0.95$. Full information on this model is presented in Table 3.9, on page 126.

In addition, the ICC values presented previously did not suffer any degradation with the addition of a fifth cluster (e.g., for Valence, ICC = 0.978, 95%CI = [0.936, 0.997], and for the PAD interaction term, ICC = 0.950, 95%CI = [0.863, 0.994]).

Despite these encouraging results, the fact that the artificial / "neutral" cluster was not automatically part of the solution issued by model-based clustering, would suggest that this cluster did not provide enough of a gain in model fit to justify the loss in parsimony. As such, this additional cluster was preserved along with the previous four only for the added benefit of facilitating inferences from experimental data, i.e., having a "neutral" stimulus category available as a baseline.

To conclude this section, the stimuli gathered for future empirical use give rise to the group averages listed in Table 3.10 (p. 126).

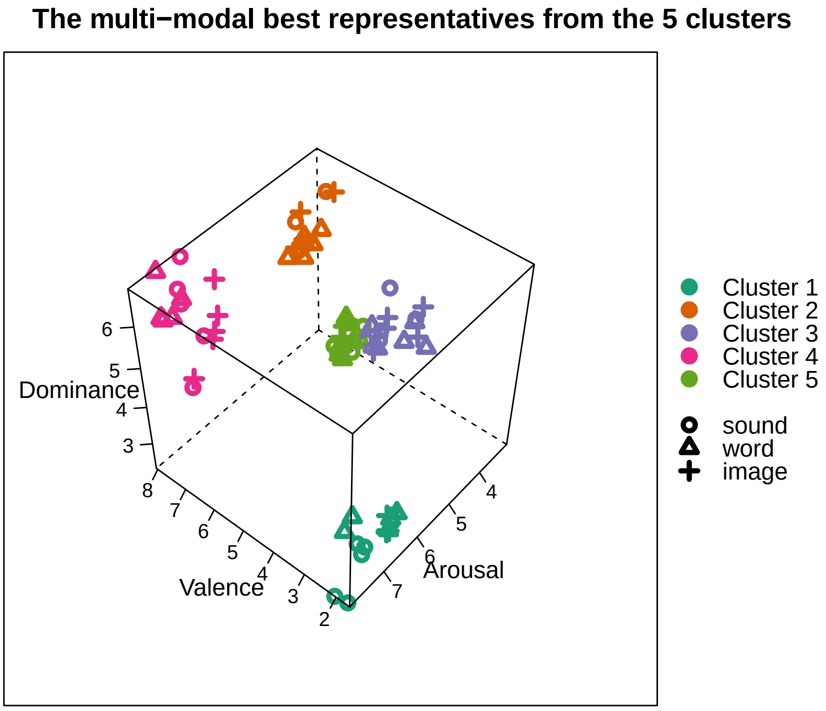


Figure 3.12: 3D plot of all the five clusters. Different stimulus modalities are depicted using different symbols, and the cluster which is closest to the middle of the cube is the one added manually. Of course, its closeness to another, data-driven cluster which is mildly negative, is apparent; however, the two are still significantly different in terms of Valence and Arousal measures.

3.3.7 Adding explanatory text / sentences

In order to be useful in research, and more easily allow the further addition of other matched modalities, we considered whether it would be beneficial to add matching affective texts to our 75 selected stimuli (i.e., 5 words, 5 images and 5 sounds within each of 5 clusters). The Affective Norms for English Text (ANET, Bradley & Lang, 2007a) were considered and tested for this purpose, i.e. providing more insight into the meaning of the current stimulus clusters, as well as some guidance on the themes that stimuli from other modalities should reflect, in case they could be added further to our selection.

The ANET represent a set of $N = 120$ short texts / sentences, which have also been normed on the PAD model. Following our previous approach, we inspected the ANET for outliers, and cases spanning more than 1 point on their 95% confidence interval. We identified and removed 7 cases due to outlying values on the Arousal dimension, and 2 and 1 cases for exceeding the admissible CI width, on the Arousal and Dominance

Table 3.9: Regression model predicting the interaction term between Valence, Arousal and Dominance, based on the cluster and modality. Coefficients are presented alongside standard errors in brackets.

	Predicting the PAD interaction
(Intercept)	124.49 (7.12)***
Cluster 1	−82.92 (8.50)***
Cluster 2	54.53 (8.50)***
Cluster 3	−33.02 (8.50)***
Cluster 4	192.24 (8.50)***
Words	7.49 (6.59)
Images	−10.69 (6.59)
R^2	0.95
Adj. R^2	0.94
N	75
Residual standard error	23.29

$^{***}p < 0.001$, $^{**}p < 0.01$, $^{*}p < 0.05$

Table 3.10: Cluster by stimulus type group averages.

No	Cluster	Type	Valence	Group Valence	Arousal	Group Arousal	Domi-nance	Group Dominance
1	Cluster 1	sound	2.04	2.05	7.47	6.92	2.56	2.87
2	Cluster 1	word	2.29	2.05	6.62	6.92	2.91	2.87
3	Cluster 1	image	1.81	2.05	6.67	6.92	3.16	2.87
4	Cluster 2	sound	7.17	7.12	4.17	4.21	6.01	5.96
5	Cluster 2	word	7.00	7.12	4.33	4.21	5.84	5.96
6	Cluster 2	image	7.19	7.12	4.13	4.21	6.02	5.96
7	Cluster 3	sound	4.48	4.26	4.29	4.26	5.04	4.96
8	Cluster 3	word	3.96	4.26	4.42	4.26	5.04	4.96
9	Cluster 3	image	4.36	4.26	4.06	4.26	4.81	4.96
10	Cluster 4	sound	7.37	7.37	7.19	7.17	6.01	5.96
11	Cluster 4	word	7.69	7.37	7.42	7.17	6.25	5.96
12	Cluster 4	image	7.06	7.37	6.89	7.17	5.61	5.96
13	Cluster 5	sound	4.87	4.95	5.01	5.03	4.89	4.96
14	Cluster 5	word	4.94	4.95	5.04	5.03	4.98	4.96
15	Cluster 5	image	5.03	4.95	5.03	5.03	5.02	4.96

dimensions, respectively.

We subsequently used the `optmatch` `R` package and `fullmatch()` routine again, to group the remaining $N = 110$ texts with our 75 pre-existing stimuli. A means for achieving this would be to specify an additional constraint for the algorithm, i.e., that the maximum number of controls for the treatment cases (i.e., `max.controls`) should be 1. This would ensure that *each* of the pre-selected stimuli (words, sounds or images) would be provided with *at least one* matching ANET sentence. Therefore, in this case, the stratum structure would reject the problematic dissipation of stimuli in groups where just one matched sentence is available for multiple actual stimuli - as happened previously when using full matching to group sounds with words.

Despite this clear benefit, such a constraint on the matching procedure would likely lead to a decrease in the overall quality of matches. Given that this analysis was carried out only for exploratory purposes and to help give some insight on the already-selected stimuli (rather than add a fully-fledged extra modality), we allowed the reuse of several sentences with different stimuli. Thus, we did not resort to the `max.controls` argument described above, and treated cases where one sentence was matched to multiple actual stimuli, by repeating that specific sentence for an appropriate number of times.

As such, the resulting stratum structure included: `1:1` matches (57.89%), more sentences matched to the same word, image or sound, of which only the sentence with the smallest Mahalanobis distance was retained (28.07%), and finally, a shared sentence for multiple words, sounds or images, where the same sentence was repeated in our results for each corresponding stimulus (14.04%). Each sentence was attributed the cluster membership of the stimulus it was matched to.

We followed up these results with a verification that the matching was as suitably efficient. If so, we would expect the ICCs to remain just as high as previously (i.e., the homogeneity of the clusters would not suffer with the addition of ANET texts). This idea was largely supported by results in Table 3.11, and the placement of data points / boxplots in Figure 3.13 (p. 128).

Table 3.11: Intra-class correlations for the five clusters including images, words, sounds, and matching sentences.

	ICC	95% CI	Groups	Average group size	Within variance	Between variance
Valence	0.92	[0.79, 0.99]	5	30.00	0.44	4.71
Arousal	0.85	[0.65, 0.98]	5	30.00	0.37	2.04
Dominance	0.78	[0.55, 0.97]	5	30.00	0.46	1.61
PAD interaction	0.87	[0.70, 0.98]	5	30.00	1689.07	11379.33

A listing of all the stimuli selected can be found in Appendix B, p. 441 (for stimulus trios), and Appendix C, p. 447 (for stimulus trios with corresponding ANET texts).

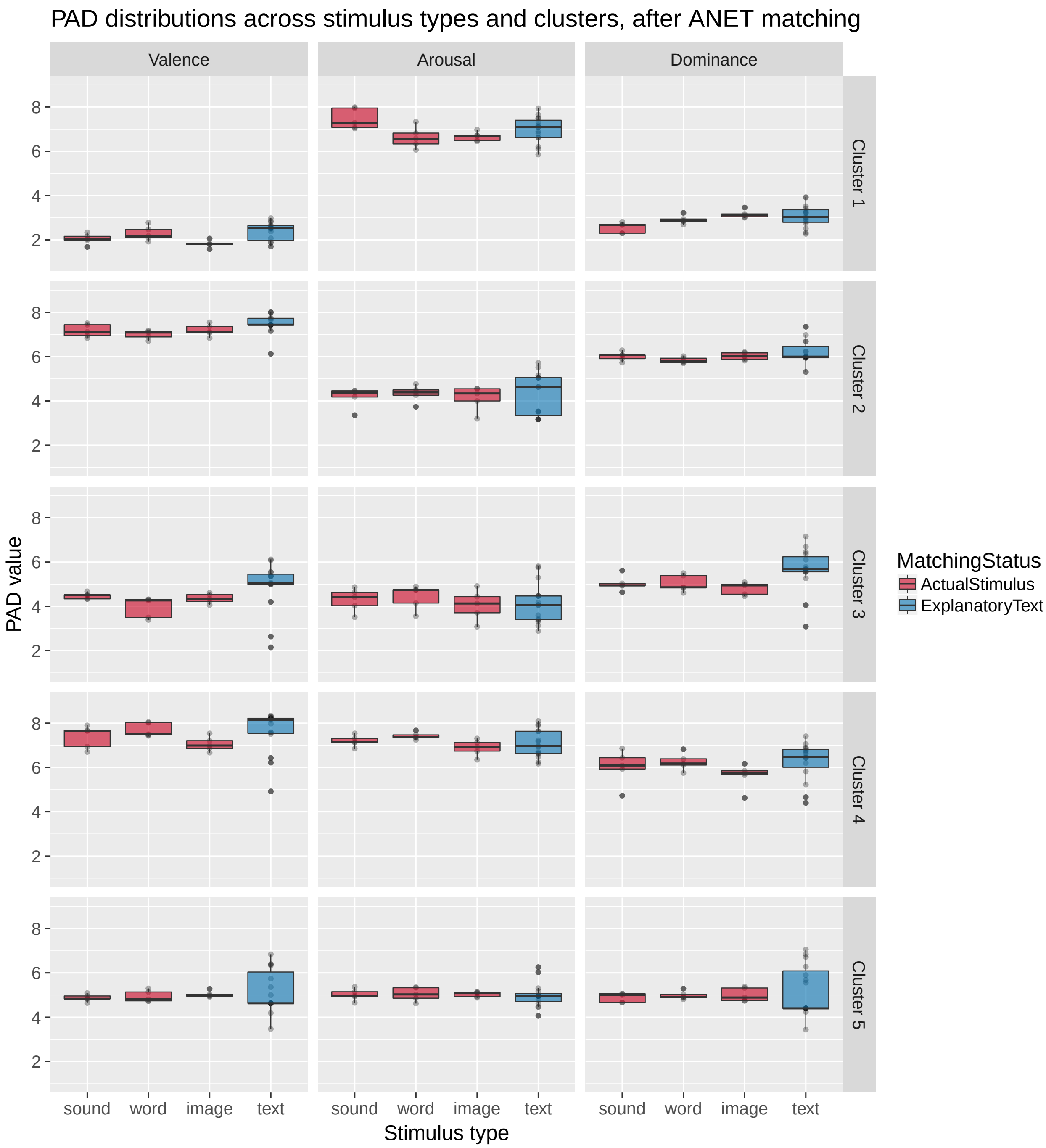


Figure 3.13: Stimuli matched across 5 clusters and 4 types: ANEW words, IADS-2 sounds, IAPS images, and ANET texts.

3.4 Discussion

3.4.1 Overview of work

In this work, we set out to create a set of matched stimuli, which could be divided into discrete groups, and used as the levels of a treatment variable in empirical research. Such a treatment variable could, for instance, be used to investigate the interactions between emotions and other cognitive processes, or the degree of similarity between multiple measures of emotion (e.g., self-reports, physiological and brain activity) etc.

In order to produce this set of stimuli, we selected three modalities / stimulus databases, which are used frequently in the literature: the IAPS (images), ANEW (words) and IADS-2 (sounds) - all of which share a common measurement model, based on Valence, Arousal and Dominance scores ranging from 1 to 9. We inspected each database before proceeding further, and excluded: duplicated cases, outliers, any missing data, and cases whose 95% confidence interval spanned over a specified criterion for judging precision (1 point, in this case).

Subsequently, stimulus matching across the different modalities was done according to two algorithms: pairwise and full matching. These led to each sound stimulus being matched to one image, and separately, to one word. In this fashion, stimulus trios were formed, and a new dataset was created where each modality was represented by an equal number of stimuli. Because of having achieved this balance, a cluster analysis then followed, which did not risk a-priori to over-sample any particular modality.

Specifically, model-based clustering was used to compute the clustering solution for this matched dataset, and based on BIC values, the most suitable number of clusters found was four. These clusters presented a "symmetrical" appearance, with two being positive (one more relaxing, and the other more arousing), and two being negative (one mildly, and one intensely so). Pending the consultation of various validation techniques and alternative clustering algorithms, this solution involving 4 clusters was shown to be robust and very well-supported.

In addition, model-based cluster analysis also outputs the probability that each case has been correctly assigned to its respective cluster. As such, we were able to rank cases within clusters in terms of their likelihood of correct membership, and sample a fixed number of such "best representatives" across all modalities and clusters. The resulting subset included five cases per modality, per cluster, therefore 60 stimuli sampled in total for further use.

Interestingly, the four-cluster, data-driven structure described above contains no obviously "neutral" cluster. As such, we considered that creating an "artificial" neutral cluster would be helpful in setting a baseline condition, against which to compare all the other clusters in empirical research. Therefore, we mimicked the structure of the other clusters extracted previously, and sampled the 5 words, sounds and images which were closest to the centre of the 3D PAD space, i.e., closest to a point in this space with the coordinates `(5, 5, 5)`, given that each PAD dimension ranges from 1 to 9.

The addition of this "neutral" cluster (with neutrality defined as closeness to the centre of the 3D space) did not have a dramatic impact on the within-cluster homogeneity / between-cluster variability, as measured by intra-class correlations. Equally, t-tests and a regression model confirmed that there are significant differences between this neutral cluster and all the others, with the exception of a comparison based on Dominance scores, between the artificial cluster and its closest neighbouring data-driven cluster, i.e., the mildly negative one.

Finally, in order to help future research with either gaining more insight into the meaning of the current clusters, or providing ideas for what stimuli to perhaps sample further and add to this current set, we have also matched ANET texts / sentences to each of the 75 stimuli (i.e., 5 clusters in total including the additional neutral one, each including 3 modalities, and each modality represented by 5 stimuli).

3.4.2 Is there a "common emotional system" based on these results?

Bradley and Lang (2000) and Redondo et al. (2008) argued that if researchers compared different, but matched stimulus modalities, any similarities in how participants respond to these modalities could be attributed to a "common emotional system". In light of the current results, it would be difficult to clearly exclude this notion, however the current evidence appears to lower the probability of that hypothesis.

On the one hand, the three modalities tested in this work (i.e., images, words, sounds) *do* partially overlap in various areas of the 3D PAD space. But for this overlap, matching the modalities into stimulus trios would have been altogether infeasible.

On the other hand, this overlap / matching process between modalities was far from perfect, as pairwise matching was possible for sounds and images, but not for sounds and words, where it was necessary to resort to full matching instead. The matching process may also be prone to bias, due to the IADS-2 sound database being relatively small, and thus not extending across the PAD space as much as it might have, or as much as the other two modalities.

Hence, overall, it is doubtful that the amount of overlap identified between these databases / modalities is sufficient to indicate the existence of a common emotional system. Furthermore, these findings are only based on self-report data, and it is unclear both how stable our stimulus trios would be, and what influence would be exerted on the matching distances used to create them, if the dependent measure were replaced by e.g., physiological or brain data, instead of self-report.

To further investigate the idea of a "common emotional system", we compared current results to our previous work (Constantinescu et al., 2016), where we clustered *only* image data from the IAPS database. Given that the current research includes the additional step of matching the various stimulus modalities *before* ever clustering them, we attribute the differences in cluster structure and placement between the *IAPS-only* solution, and the clustering solution for images as part of *matched trios*, to the matching process itself.

More specifically, using colour-coded clusters in the background, and specially emphasised cluster representatives in the foreground, we visualised and compared the cluster structure derived based on IAPS data only, to the image counterparts from our matched trios. It emerged that not only did the number of clusters change between these conditions (i.e., five in the IAPS-only solution, and four in the matched data, to which a fifth was artificially added), but also the general placement of the clusters and their representatives shifted considerably after the matching process (which, to reiterate, was dictated by sounds). Figure 3.14 (page 132) shows this drift.

Generally speaking, the fact that matching images to two other modalities changed the clustering structure to this extent, indicates that there is probably little support for a "common emotional system", despite, interestingly, all three modalities (as well as ANET texts) sharing the distinctive quadratic trend / "U"-shaped relationship between Valence and Arousal (a feature also detected by Bradley & Lang, 2000, and discussed previously in Figure 3.4, page 111).

Nonetheless, the matching process proved to be a highly useful tool, or starting point, in the endeavour to investigate any commonalities (or perhaps, differences) between affective modalities. For instance, this process would allow for a much safer comparison between elicitation methods, as the influence of confounding variations in Valence, Arousal and Dominance on dependent measures, would be limited.

Despite their clear utility, it is very uncommon to encounter such matching algorithms in research using multiple stimulus databases simultaneously. The reason for this may relate to the obscurity of these computational methods, and the general preference for 'ad-hoc' / manual stimulus sampling strategies. A search on Google Scholar using the query: ["`IAPS`", "`ANEW`", "`IADS`"] on April 27, 2017, yielded 183 results, of which only 3 were considered directly relevant to the issue of sampling / matching stimuli from multiple modalities / databases, for use in empirical research.

For instance, Lepping, Atchley, and Savage (2016) attempted to match IADS sounds to music clips, in order to compare these categories more meaningfully in empirical research. The matching of stimuli in this study was done by means of an analysis of variance (two conditions: music vs. IADS sounds, by two Valence categories: positive vs. negative), and it was considered that the two conditions would be matched if the analysis of variance was non-significant. However, this method is fairly rudimentary, and

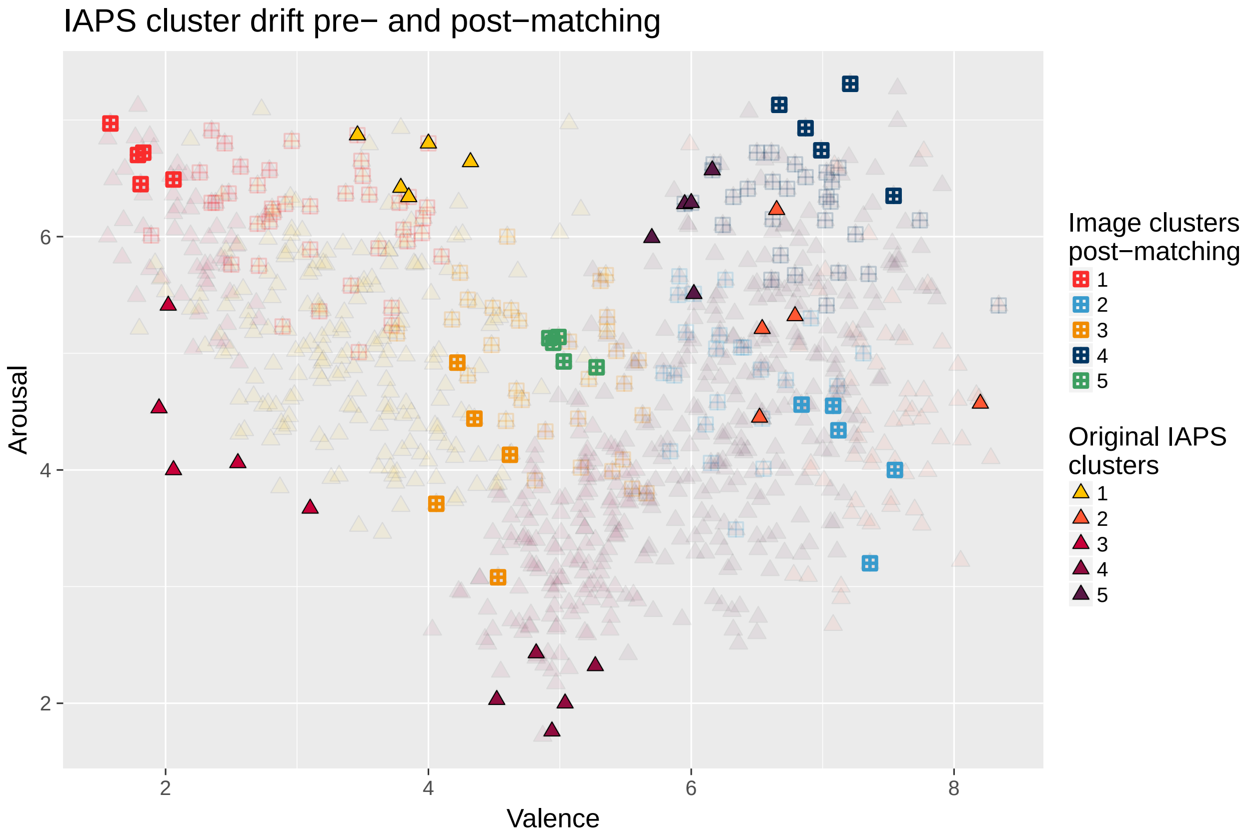


Figure 3.14: Image cluster drift is shown pre- and post- stimulus matching using different symbols: triangles for IAPS images clustered independently of any matching process, in Constantinescu et al. (2016), and crossed squares for the remaining images clustered after the matching process. In the background, the full data in each cluster is displayed using transparent points, whereas the solid colour symbols in the foreground represent the best representatives in each cluster. Interestingly, the images (and their best representatives) remaining after the matching process, tend to present higher Arousal in this 2D space relative to the initial set of IAPS data, which shows more extensive coverage of this space. This suggests that matching images to sounds and words lead to a bias in how images were sampled, and is problematic for asserting the existence of a "common emotional system".

does not ensure that the stimuli are matched one by one, particularly since null results are typically difficult to interpret (De Graaf & Sack, 2011; Ferguson & Heene, 2012). Hence there is still room for improvement, and we propose that a matching algorithm such as that provided by the `optmatch` package in `R` would provide the much-needed assistance in this area.

In addition, Larsen et al. (2003) sampled stimuli from all three databases (IAPS, ANEW and IADS), in order to investigate the effect of stimulus Valence on facial muscle activity. To achieve this, the Valence dimension from each database was cut into 11 discrete categories, and 6 stimuli were sampled from each, thus leading to a number of

66 stimuli sampled per modality / database. Due to this method of sampling, the stimuli were *not* actually matched, leaving room for error when comparing muscle activity patterns between modalities. This is apparent in Figure 3.15, on page 134, where we have contrasted our matched distributions to Larsen et al.'s (2003)[6]. As is easy to see, our matching method has clear benefits in terms of limiting confounds in research.

After this initial sampling of 66 stimuli per database, Larsen et al. (2003) then used a local sample to re-rate them in terms of their Valence (using unipolar scales, rather than the usual bipolar SAM scales, to measure the extent to which stimuli are mostly positive, mostly negative, or ambivalent/neutral). Based on this data collection, the authors further aggregated their stimuli into five categories: Very Negative, Mildly Negative, Neutral, Mildly Positive, and Very Positive. Unfortunately, the codes for which stimuli were aggregated in which category, were not made available.

However, using the only information that was indeed reported by these authors, Figure 3.16 (page 135) illustrates the 66 stimuli sampled from each database, alongside the best representatives from our own clusters. While the sampling of cases by Larsen et al. (2003) does extend nicely across the stimulus distributions and does not seem to ignore any particular area within the PAD 3D space, it is also true that without a data-driven approach for matching and grouping stimuli, effort may be unnecessarily expended for testing stimuli which ultimately do not represent key areas of interest in the PAD space (i.e., the cluster best representatives marked in the same plot). For this reason, as well as mismatch in stimulus distributions discussed previously, we would recommend that matching and cluster analysis algorithms be used instead of this type of approach.

Finally, in their affective judgement task, Norris, Larsen, Crawford, and Cacioppo (2011) used the same choice of modalities, sampling strategy, and stimuli as Larsen et al. (2003), so our same comments would apply.

3.4.3 Is it worthwhile to use cluster analysis to sample stimuli? If so, what do the clusters mean?

As we have argued previously (Constantinescu et al., 2016), (model-based) cluster analysis is an excellent alternative to "manual" stimulus sampling methods, as it ensures that stimulus groups are as different as possible, and also offers important information on which cluster members are likely to have been correctly classified, vs. which are ambiguous in terms of their cluster membership (i.e., in the case of stimuli places around the borders of two neighbouring clusters). This feature can be exploited in order to extract best representatives from each cluster, i.e., those stimuli with the highest likelihood to have been correctly classified.

[6] Which was possible given that these authors reported the stimulus codes they used.

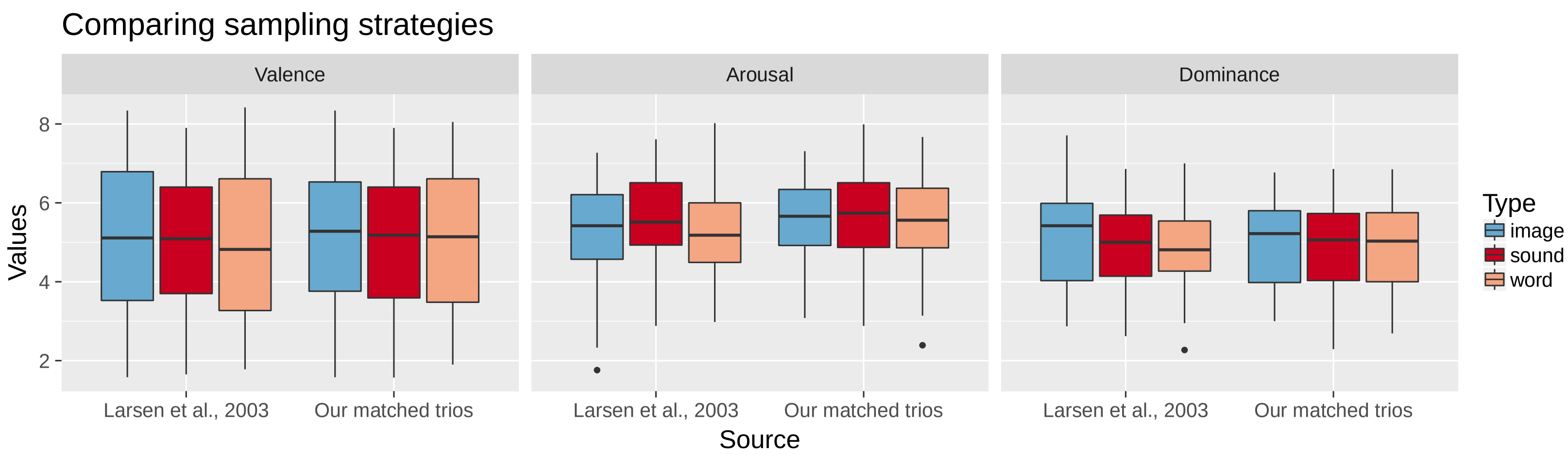


Figure 3.15: Presenting the stimulus distributions sampled by Larsen et al. (2003) in parallel with our own, shows that considerable variation remains between modalities, if they are not specifically matched before use in research. This can be problematic if the intention is to compare these modalities in terms of e.g., the pattern of facial muscle activity, given that PAD variations can then act as confounds.

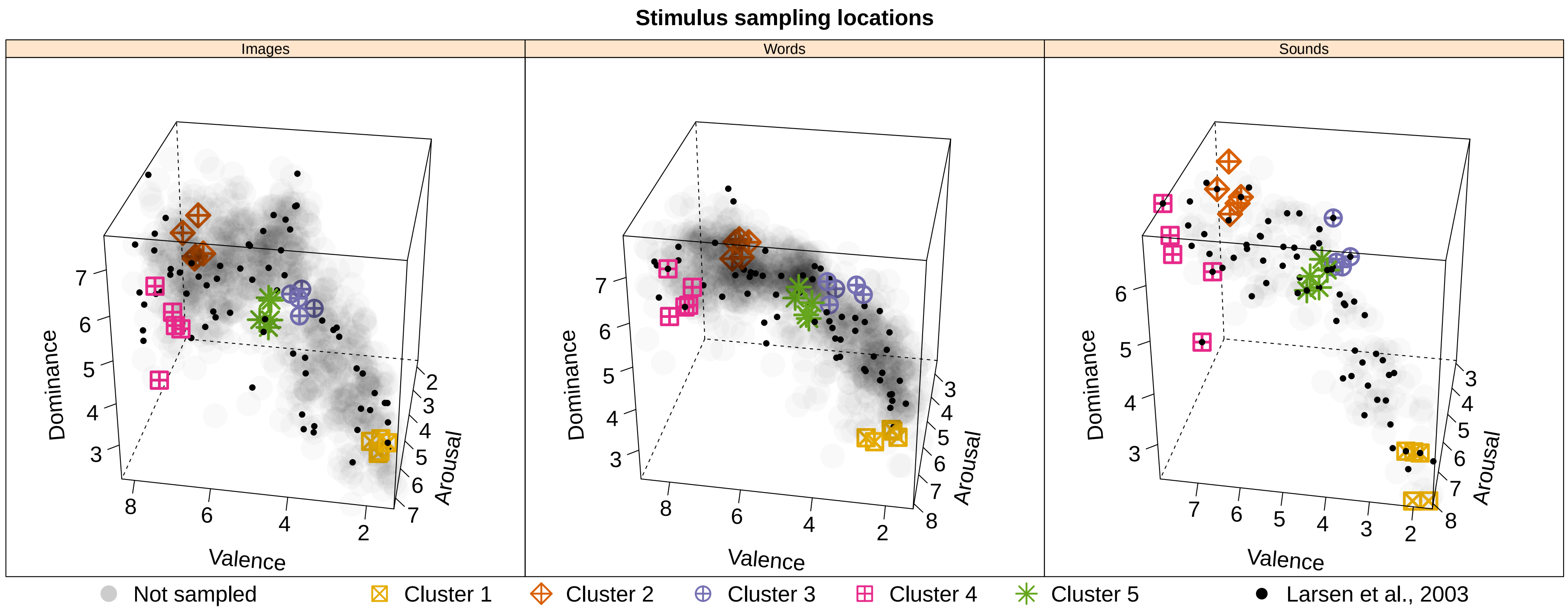


Figure 3.16: Against a faded background representing the full distributions of IAPS, ANEW and IADS-2 stimuli, in this plot we emphasise the stimuli sampled by Larsen et al. (2003) (using simple black dots), as well as our own best representatives, from each of our five stimulus clusters. While Larsen et al. (2003) sample stimuli which seem to nicely represent the entire distribution of points from each cube, it becomes clear by looking at the placement of best representatives from our own clusters, that time and effort will be spent in this way on testing stimuli which do not maximise between-group differences a-priori. The same authors unfortunately do not report how they subsequently grouped their stimuli into five categories.

Using this method jointly with the matching of separate stimulus modalities, this becomes a powerful tool for identifying an appropriate number of stimulus groups and how stimuli should be divided between them. Our current work indicates that this method can: create distinct and intuitively meaningful clusters; it can be adapted to include multiple modalities of stimuli; reduces the effort to sample stimuli from PAD areas that might prove to be less interesting empirically (as per the discussion above, regarding the Larsen et al., 2003 study); it can (to a certain extent[7]) adaptively accommodate the extraction of varying numbers of best representatives from each cluster; and can take into account all PAD dimensions simultaneously to create stimulus groupings - unlike manual methods.

As mentioned earlier, failing to use this methodology and resorting instead to simpler sampling strategies, can result in unwanted consequences. For instance, in work by Miles and Johnston (2007), Arousal and Dominance were not taken into account when sampling stimuli, therefore allowing the possibility of confounds in results. Also, in research by Keil et al. (2007), the chosen neutral and positive stimulus groups overlap to a large extent, with precise consequences for power and interpretability remaining unknown. Both these issues could easily be overcome using model-based cluster analysis to sample stimuli.

The four clusters we uncovered in the mixed-modality data are symmetrical in structure, and represent intuitively meaningful content: for the mildly negative and moderately arousing cluster, an image of rain, sound of a clock, and the word "`solemn`" were included as best representatives; for the very negative and highly arousing cluster, best representatives were, e.g., the image of a dead person, the sound of an attack occurring, and the word "`drown`"); the positive and fairly unarousing / relaxing cluster involved e.g., the image of flowers, the sound of a harp, and the word "`politeness`") as best representatives; and finally, in the very positive and arousing / exciting cluster, e.g., the image of a rollercoaster, the sound of bongos playing, and the word "`sexy`" were incorporated. Interestingly, these results confirm the findings of Stevenson et al. (2007), and Stevenson and James (2008) in spirit: the PAD model complements the discrete view on emotions, but does not reproduce it.

We believe that the concept of "neutrality" merits a somewhat more detailed discussion, however, given that cluster analysis has proven to be a highly useful tool for investigating this in such stimulus databases. In our previous work (Constantinescu et al., 2016) based on clustering images only, we discovered that the cluster which best reflected the idea of neutrality was not situated, as one might expect, in the centre of the 3D PAD space (i.e., around a score of 5 on each of the PAD scales, as measured

[7] Unlike the IAPS classification work, this feature is severely limited due to the small N of the matched data, which, in turn, is attributable to the small N of the IADS-2 sound database.

on 9-point Likert scales). Instead, the IAPS neutral cluster was defined by a centroid close to a score of 5 only on the Valence scale, whereas in terms of Arousal, it presented low scores, and high scores on Dominance. This interpretation of neutrality is therefore consistent with reassuring, and unchallenging content (e.g., such as a spoon or basket), which in fact seems sensible.

However, when matching IAPS images to IADS-2 sounds, and to ANEW words, the neutral cluster collapsed altogether, leaving a symmetrical structure of just four clusters which were otherwise similar in spirit to those found previously: two negative clusters (one mildly, and one more intensely so), and two positive clusters (one more serene, and one more arousing).

The reason for the disappearance of the fifth neutral cluster is related to the matching process itself: the sound data available for matching was the major limiting factor for the entire process, and led to the selection of images (and words) being biased according to whatever sounds were present in the IADS-2 database. Given that sounds were overall more arousing than either of the other modalities (which is clearly visible in Figures 3.3 and 3.4, on pages 110-111), the images (and any suitable words) which could have populated the medium-Valence, low-Arousal (and high-Dominance) neutral cluster simply could not be sampled, since no corresponding sounds exist with such a low level of Arousal. It is difficult to say whether this is a fact that is intrinsically related to the sound modality itself, or the IADS-2 database simply does not contain enough stimuli to adequately represent the PAD space, in the same way that IAPS images (or ANEW words) do.

Be that as it may, in order to still benefit from the existence of a neutral cluster (which is a methodologically desirable feature for a stimulus pool), while still satisfying the constraints imposed by the distribution of sounds on the matching process, we crafted an "artificial" neutral cluster. In this case, this mixed-modality, neutral cluster did not (and could not) include medium-Valence, low-Arousal and high-Dominance stimuli, but rather, in accordance with the types of sounds available, more naturally accommodated medium scores on all three PAD dimensions. This cluster was most similar to (but still significantly different from) the mildly-negative stimulus cluster already issued by model-based clustering.

Based on the stimulus norms made available with the databases, cases such as: the sound of a pig (norms: `4.64, 4.93, 5.00`), the word "`army`" (norms: `4.72, 5.03, 5.03` on the PAD model), and the image of a battleship (norms: `4.95, 5.09, 4.89`), were sampled as part of the artificial *neutral* cluster. While these stimuli may not necessarily strike the reader as being obviously neutral, their PAD coordinates place them towards the very centre of PAD 3D space. Shortly, we will also briefly discuss cultural differences as a limitation for this study, given that the norms used in this analysis were gathered from American samples.

3.4.4 Limitations

A few limitations and caveats are in order, after having discussed our method.

Firstly, the major drawback to the results presented here concerns the IADS-2 database, which, compared to the other two modalities used, is very small. This imposed significant difficulties in adequately sampling the mixed-modality PAD space, given that no sounds existed to cover areas of the PAD space, where images and (sometimes) sounds still extended. This introduced bias into which images and words were sampled as matches for the sounds available - with perhaps the most serious consequence being the disappearance of the “neutral” cluster found in our previous work on the IAPS only (i.e., a cluster which now occupied a PAD space devoid of sounds). With the addition of more sound stimuli into the IADS-2 database, this issue would probably be alleviated; for now, however, it was overcome by “manually” creating a neutral cluster, comprising of stimuli which were closes to the centre of the PAD cubic space.

Secondly, the database norms were based on an American sample, which might possibly explain why e.g., a word such as “army” or the image of a battleship are close to neutrality, or, the aforementioned centre of PAD space. It is unclear how the matching process and clustering results would differ, if used on data from a different culture (e.g., if obtaining data used by Redondo et al., 2007 when re-norming the ANEW on a Spanish sample).

Finally, a word of caution is perhaps worth mentioning with respect to statistical power. While the clustering method does ensure that stimuli are as different as possible between clusters (thus supporting a boost in effect sizes), it is also limited by the small sample size of the IAPS-2 sound database, and hence, the small number of trios fed into the clustering procedure. As such, the number of best representatives sampled per modality, per cluster (i.e., the number of usable trials in research), will be relatively small - and consequences for statistical power will vary across research topics and designs.

3.4.5 Future directions

In future work, we aim to further add other modalities to this matched set of stimuli, particularly film clips and virtual environments. We believe that preparing such a pool of mixed stimuli will prove useful in the study of how different modalities affect the process of emotion elicitation in controlled environments. As support in the process of adding new stimulus modalities, we believe that the ANET texts added to our matched data will provide additional insight into what type of content to search for, in the case of films and virtual environments.

3.5 Conclusions

In this study, we present a method for sampling stimuli from different modalities (i.e., sounds, words and images), and matching them on a case by case basis, using their Valence, Arousal and Dominance scores. Following the creation of a matched dataset, where each sound has a corresponding image and word, we created a clustering solution which classified these matched stimuli into four clusters: two types of negative, and two types of positive clusters. The absence of a naturally-emerging neutral cluster was compensated by artificially crafting one such cluster, thus leading to a total of five clusters each containing 15 stimuli (i.e., 5 words, 5 sounds and 5 images), or 75 stimuli overall, ready for use in empirical research. Our method limits confounds by matching the stimulus modalities, easily takes into account all 3 PAD dimensions (unlike manual sampling methods), creates optimal and parsimonious stimulus groupings, and identifies intuitively meaningful clusters, best described by their best representatives, or the cases with the highest likelihood of having been correctly assigned to their respective cluster.

Part III

Comparing five different types of emotion eliciting media: words, sounds, images, film clips, and virtual environments

Introduction

In affective science, a high amount of variation is present across research methods and instruments, as discussed previously in the introductory chapter (Part I, from p. 29). In this context, the importance of consistency cannot be overstated, with the matching and comparison of multiple elicitation methods and measurement instruments holding promise for a more profound and coherent understanding of this area.

However, the benefits of such comparison work run deeper than just improving coherence in the field, particularly when contrasting various elicitation methods can *also* serve to investigate the existence of a general affective system - one responsible for managing responses to affective stimuli, regardless of their modality. Should such a system exist, it should prove relatively easy to match stimuli across different modalities, based on their PAD (Valence / Pleasure, Arousal and Dominance) scores - which are often considered as the most basic level of evaluating and describing emotional phenomena (Mehrabian, 1995).

Furthermore, work on stimulus comparison and cross-modality matching may also have uses for limiting the influence of the PAD dimensions as covariates, at times when different outcomes are under assessment, e.g., the level of engagement induced by stimuli. In other words, if stimuli have been matched on PAD dimensions, we can more easily ask whether they *still* differ in terms of other dimensions, e.g., how engaging they seem.

This last question is perhaps more important than it would appear at first glance. Ideally, we believe that "effective" lab stimuli and tasks should induce high levels of engagement. This is because, in contrast, non-engaging, "artificial" stimuli would encourage results with less relevance for the "real world ecologies" where emotional phenomena were designed to operate (Brunswik, 1955; Hammond, 1948, also see Brunswik's model, in Figure 1.5, Section 1.4.2, p. 55).

In addition, for the study of emotions, "artificial" tasks might also increase the frequency of over-interpreted and consciously considered participant responses - which are ultimately less valuable in affective research, because they tend to reflect cognitive, instead of affective processes. Not only this, but a less engaging lab setup is also likely to induce fatigue / boredom sooner, which again could damage the validity of participant responses.

For all these reasons, a comparison of multiple methods of emotion elicitation serves the essential goals of: trying to discern whether a low-level, central emotional processing system exists which is based on the PAD model, and if so, whether a variety of such elicitation methods can be matched on this model. Further, we also investigate which of these elicitation methods appears to be more engaging for participants, in the hope that such stimuli can lead to lab data with higher relevance for the outside world.

Matching and comparing commonly used stimuli for eliciting emotions

To achieve a comparison between affective media, where available, large stimulus databases were used, i.e. the IAPS (for images, Lang et al., 1999), ANEW (words - for which an older version of the database was used as it also included word frequencies, Bradley & Lang, 1999a) and IADS-2 (sounds, Bradley & Lang, 2007b). These have been discussed at length in Section 1.3, and the Introductions of Chapter 2, and Chapter 3.

For the purposes of this research, we also included two further modalities: film clips and virtual environments (VEs). These stimuli were selected on an individual basis either due to the unsuitability of any already available databases (discussed in more detail below), or simply due to the novelty of the technique (for VEs), which meant such databases are not yet available. Films and VEs were included in this research due to presumably inducing higher levels of engagement and *presence*, relative to the other modalities investigated.

Presence has been discussed in relation to VEs in particular, and can be defined as a feeling of actually existing and operating within a highly engaging environment (Baños et al., 2004; Lessiter, Freeman, Keogh, & Davidoff, 2001; McMahan, 2003; Schubert, Friedmann, & Regenbrecht, 2001)[8]. Arguably, generating a feeling of presence should encourage participants to behave as they normally would outside testing conditions, and enable emotional stimuli to function as a proxy for real-life. Hence, after matching the various modalities investigated here on the PAD model, a secondary aim was to track whether any of them may be particularly suitable for inducing presence.

[8] Importantly, we adopted the recommendation of Slater (1999); Slater and Wilbur (1997) (also implemented by Diemer, Alpers, Peperkorn, Shiban, & Mühlberger, 2015), and have been referring to technological manipulations as manipulations of *immersion*, rather than of *presence* - which instead is the subjective feeling arising from the use of such technologies.

Films

Films are some of the most commonly used and most effective types of stimuli for inducing emotional states, particularly in controlled environments (Westermann et al., 1996). Despite this, Bartolini (2011) claims that only three film clip validation studies were conducted prior to 2011.

This may be related to the fact that it is relatively common to create one's own selection of film stimuli, rather than opt for extant stimulus databases, which are more likely to be associated with larger-scale validation work. Indeed, we explored the list of articles used by Lench et al. (2011) in their meta-analysis (for details see Section 1.3.2, p. 44) - which included 106 studies where films were employed to elicit emotions. We explored a number of these ($N = 42$) in more detail, and found that using and pretesting new films is a relatively common practise in emotion elicitation studies, compared to using pre-established film databases. Nonetheless, several film databases do exist, such as those constructed by J. J. Gross and Levenson (1995), A. Schaefer et al. (2010) and Carvalho et al. (2012).

To begin with, perhaps the best known film database was created by J. J. Gross and Levenson (1995), and included 16 clips. It is based on a discrete emotions perspective, however, with 2 clips included for each emotion considered. The database offers instructions for recreating the film clips whenever these were taken from commercial films, or, for the subset that are non-commercial, the database includes them directly. However, the fidelity of these films is outdated relative to modern standards, and some appear to lack sound, which could act as confound variable.

Subsequently, A. Schaefer et al. (2010) have also collected an impressive database of films for emotion elicitation, including normative ratings, as well as a classification of stimuli according to both the dimensional view (with Valence and Arousal ratings included), or according to the discrete emotions view. However, this research was aimed at a Belgian, French-speaking population, which makes this database risky for use on an English native-speaking population. This is also suggested by the findings of Roberts and Levenson (2006), who discovered that for some cultures, emotional responsivity is higher when the characters in the films represent the same culture.

Finally, Carvalho et al. (2012) have also collected a smaller film database based on the PAD model. However, these are unimodal (solely visual), which may limit comparisons between studies. Hence, due to the drawbacks presented for all the previously mentioned film databases, as well as the frequency within the field of pretesting new films not included in databases, we decided to search for our own film clips to use in the current research.

Virtual Reality (VR)

The idea that real environments could be simulated on computers for research purposes is not new, e.g., "micro-worlds" (or game-like environments anticipating later virtual environments) were proposed as an intermediary testing medium between the research lab and the field (Brehmer & Dörner, 1993). There is in fact empirical evidence suggesting that individuals behave similarly to real-life when immersed into virtual environments (VEs), in terms of interpersonal behaviours and empathic reactions to characters in need (Gillath, McCall, Shaver, & Blascovich, 2008), or emotional reactions to (virtual) food when diagnosed with an eating disorder (Gorini, Griez, Petrova, & Riva, 2010).

In this context, Virtual Reality (VR) is receiving increasingly more attention due to its applications for research (Fox et al., 2009; Loomis et al., 1999). For instance, Pallavicini et al. (2013) and Parsons (2015) discuss the uses of VR for boosting 'ecological validity'[9] and experimental control in clinical, affective and social research settings. Parsons, Gaggioli, and Riva (2017) also promote its usefulness, by contrasting VR to other typically used, and more simplistic stimuli (e.g., stills of facial expressions, or IAPS images, affective texts etc.) - which may not be representative of more complex real-life situations and tasks. The same authors propose that research testing conditions should be more similar to real-life events which blend multimodal, dynamic and contextual information. They believe VR is uniquely suited to emulate these, because it does not require that experimental control be sacrificed in the name of realism, can be embedded within wider narrative contexts, enables the integration of multifaceted measurement strategies (i.e., behaviour, self-report and physiological measurements), and finally, even allows the investigation of impossible situations (e.g., embodying a self of another race etc.).

VR has also pervaded areas of mood/emotion elicitation, and emotional response modulation. For instance, Felnhofer et al. (2015) successfully used virtual parks to induce joy, anger and anxiety, and less successfully - sadness and boredom. Serrano, Botella, Baños, and Alcañiz (2013) also used VR to induce relaxation, whereas Chirico, Yaden, Riva, and Gaggioli (2016) relied on VR for inducing awe. Moreover, VR has even been used to treat affective disorders (e.g., phobias or food disorders, Gorini et al., 2010; Parsons & Rizzo, 2008).

Based on these findings and the relevance of VR for eliciting emotions, we explored various possibilities for interacting with online virtual worlds, so that the most suitable platform could be chosen for our research. Neuro-VR[10] was investigated, as well as High-

[9] Or rather, representative experimental design, if using Brunswik (1955)'s original terminology.
[10] http://www.neurovr2.org/

Fidelity[11], Blue Mars[12], and Second Life[13]. At the time of the search, Neuro-VR seemed to offer a rather limited range of downloadable environments, despite being able to output images to a head-mounted display (HMD). High-Fidelity, although arguably the most promising in terms of the realism of the characters (programmed to display surprisingly realistic facial movements and intonations), was still in its development stage. Blue Mars, while boasting very high rendering quality (superior to all the other worlds investigated at the time) did not appear to be compatible with HMDs and, precisely because of its level of detail, risked being very slow-moving depending on bandwidth and processing speed.

Hence, we placed more emphasis on our investigation of Second Life (SL), and extracted review data for 215 online virtual worlds/platforms from the website: http://reviews.virtualworld.com/reviews/[14]. According to the overall scores given to each world reviewed, Second Life was classed among the best 10% of the virtual worlds tested, in terms of "Price to value", "Game quality", "Member base" and "Ease of use".

Additionally, work posted by Dr. Sarah Smith-Robbins[15] on her personal blog (live at the time of the search, April 12, 2014) also revealed that among the virtual worlds considered, Second Life was the leading option when considering various criteria, e.g., if the environment is *stigmergic* (i.e., preserves traces of a user's activity, so that other users can build on it), supports visual user to user communication, and collaborative relationships between users, or between users and the environment etc.

For these reasons, as well as the existence of a viewer program (CtrlAltStudio) capable of outputting visual information to an Head-Mounted Display (the Oculus Rift), Second Life was the preferred solution for introducing participants into (immersive) virtual environments.

Overview of studies

Due to the complexity of the research aims discussed previously, this investigation was divided into three parts/studies: the first included the matching and comparison of words, sounds, images and film clips; the second focused on matching and comparing film clips to non-immersive VEs - where any VEs used were viewed by participants on a large computer screen; and finally, in the third study, we verified whether or not the use of immersive equipment (Head-Mounted Displays, HMDs) can affect previous results.

[11] https://highfidelity.com/
[12] http://www.bluemarsonline.com/
[13] http://secondlife.com/
[14] Still live at the time of the search, August 5, 2014.
[15] Senior director of emerging technologies at Kelley Executive Partners at Indiana University. Search conducted on March 10, 2014.

Chapter 4

Study 3A: Comparing words, sounds, images and films

4.1 Aims

As part of a set of three studies, Study 3A opens our line of inquiry with several interconnected objectives: firstly, we will investigate how the data derived from an Edinburgh student sample might differ from IAPS / ANEW / IADS-2 norms obtained from an American student sample (the latter of which were used in our process of selecting stimuli for research use). This will serve as an indication for how sound an undertaking it is to sample affective stimuli based on norms from a foreign culture, and for how generalisable our results will be.

Secondly, we will select and match a subset (i.e., one third) of the 75 YouTube clips we gathered, to the other three modalities in terms of PAD ratings (represented by 25 stimuli each), and thus stimulus quartets will be created, where each will include: 1 word, 1 sound, 1 image and 1 film clip. This will serve the purpose of controlling cross-modality variations in PAD values, and will be particularly useful when, for instance, predicting other characteristics such as engagement.

After the matching process, we will check whether the AffectButton and SAM scales provide similar information on the stimuli, i.e., if there are correlations between a PAD dimension as measured with the SAM, vs. with the AffectButton. If any important degree of convergence between these measures is found, this will testify to the validity of the AffectButton.

Subsequently, we will also verify whether the cluster structure we initially devised based on stimulus norms (i.e., 4 'organic' clusters, and 1 neutral cluster artificially added after the fact) in the previous chapter, re-emerges on a sample from Edinburgh, UK. For this, we will re-run a previous model-based cluster analysis, as well as predict stimulus categories (or clusters) from participant ratings.

In addition, we will also check for covariates which were measured at the beginning of the study, and which may affect PAD ratings for our stimuli, e.g., participant or stimulus characteristics.

Finally, we will investigate how some of the stimuli varied on measures of presence and engagement - done in preparation for Studies 3B and 3C, where VR will be introduced as a modality as well, to compare against those assessed currently.

4.2 Method

4.2.1 Participants

Sixty participants were recruited via the University of Edinburgh 'My Career Hub' service[1], with some of their characteristics described in Table 4.1 (page 150). On the whole, the sample consisted of: 40.00% Western Europeans, 26.67% Eastern Europeans, 21.67% Asian participants, 8.33% North Americans, 1.67% South Americans, and another 1.67% participants from the Middle East. Also, across the entire sample, the average age was 23.62 years (SD = 5.58 years, range = 18 - 46). Participants were remunerated £15 in exchange for completing the study.

Table 4.1: Sample description in terms of nationality, gender, and average age.

	Nationality	Gender	N	Average age
1	Eastern Europe	Female	8	22.12
2	Eastern Europe	Male	8	21.00
3	Far East	Female	8	26.63
4	Far East	Male	5	24.20
5	Middle East	Female	1	25.00
6	North America	Female	3	25.67
7	North America	Male	2	24.00
8	South America	Female	1	36.00
9	Western Europe	Female	9	20.33
10	Western Europe	Male	15	24.60

An adapted version of the exclusion criteria from WAIS-III (Wechsler, 1997) was first displayed to participants so that they could opt out of the experiment if any criteria were applicable to them, i.e.:

`Inclusion criteria: There are no pre-requisite skills for participating in`

[1] https://www.ed.ac.uk/careers/mycareerhub

```
this study, but in order to participate, none of the following can apply
to you:
- Colour-blindness
- Uncorrected hearing loss / visual impairment
- Current treatment for alcohol or drug dependence
- Heavy alcohol use
- Seeing a doctor/other professional for memory problems / problems with
  thinking
- A condition that would prevent arm/hand movement when using the keyboard
  or mouse
- Any period of unconsciousness greater than 5 minutes
- Head injury resulting in hospitalisation for more than 24h
- Currently taking antidepressant, anti-anxiety, or anti-psychotic
  medication
- Medical condition that could potentially affect study responses, such
  as:
    - stroke
    - epilepsy
    - brain surgery
    - encephalitis
    - meningitis
    - bipolar disorder
    - alexithymia
    - depression.
```

4.2.2 Materials

4.2.2.1 Words, sounds and images

These stimuli were selected based on our previous work, outlined in Chapter 2, p. 65 (also Constantinescu et al., 2016), and Chapter 3, p. 99. Full listings of these stimuli are also presented in Appendix B (p. 441) and Appendix C (p. 447).

4.2.2.2 Film clips

In order to add film clips to the set of previously matched stimuli, it was decided that the most suitable option would be to search for video material on YouTube (primarily), while enabling the Creative Commons Licence search filter. This granted us permission to freely use and modify the video material to suit our research purposes. However, in the case of erotic material (which was not available on YouTube, at least under the type of license we sought), the search was moved to other websites which can be disclosed by request. These were not under a Creative Commons license, however it was decided that their use would be permissible, and would fall under a Fair Use license instead (used for research and non-profit purposes). Every effort was made to find good quality and relevant material to match the previous stimuli (words, sounds, images), while enlisting the help of matched ANET texts as well, included to direct our search (see Appendix C, p. 447).

Overall, finding suitable stimuli presented unexpected challenges due to: the relatively small number of videos posted under a Creative Commons licence, rather than a Standard YouTube licence; the occasionally misleading titles posted for videos; frequently finding the material either too mild or too intense (i.e., either impractical or unethical), relative to the other modalities already sampled; rendition issues (shaky camera angles, image out of focus, pixelated frames etc.); short-listed and suitable videos being removed from YouTube before they could be downloaded for this study.

Seventy five film clips were selected (see Appendix Section D.2 for a list), with each of the 5 stimulus clusters (neutral, mildly negative, intensely negative, positive and serene, positive and exciting) being assigned 15 film clips, but only 5 words, 5 sounds, and 5 images. This ratio of `3:1` between films and every other modality was implemented as a buffer, and in order to ensure that there were enough film clips to be matched on PAD values as closely as possible to the previously clustered stimuli. As a result, we searched for 3 times more films, in order to ultimately use just 1/3 in analyses (empirically, those shown to be the closest matches for the other stimuli), and filter out the other 2/3.

Clips of approximately 40 seconds were sought within each of these 75 films, which varied widely in terms of their overall duration. Extracting precisely 40 seconds was not always possible, due to our intention to preserve coherence and meaning (e.g., avoiding interruption of sentences or actions performed in the film, so that the clip would make sense and could stand alone). Thus, the average clip duration was in fact 40.48s (40.40s for a 5% trimmed mean), with a range of 3.12s, stretching from a minimum of 39.21s to a maximum 42.33s, and with a standard deviation of 0.59s.

For replicability, Appendix D.2 contains the exact frames (and rough time-points) where the original films were cut in order to form our stimuli. The open-source software used for video editing was Avidemux[2]. The settings used within the program were for the audio stream: `MP3 (lame)`, for video: `MPEG-4 ASP (Xvid)`, and with the `.mp4` format selected. We also used the open-source **`ffmpeg`**[3] software on Ubuntu to extract details about these film clips, i.e., frame rate / number of frames per second (FPS), bandwidth[4],

[2] https://avidemux.en.softonic.com/

[3] https://ffmpeg.org/

[4] A measure for the "richness" of visual information conveyed by the film, equivalent to the number of bits processed in a unit of time.

and resolution, for later use as covariates.

It is worth noting that some of the clips have background music, which can also be used as an independent form of emotion elicitation. Selecting film clips also based on this criterion (i.e. avoiding those which also contain music, rather than just sounds relevant to what is being filmed) would have restricted the search further, and was not feasible. As such, 22 of the 75 clips (i.e., 29.33%) contain such background music, including: 2 assigned to the neutral cluster (of a total of 15 films per cluster), 1 assigned to the mildly negative cluster, none in the intensely negative cluster, 9 in the positive and exciting cluster, and 10 in the positive and serene cluster. This measure will also function as a covariate in subsequent analyses, to verify whether it may bias ratings.

All the stimuli used - including information about modality and cluster membership - are presented below, in Table 4.2).

Table 4.2: Full stimulus list including the blocks of words, sounds, images and film clips, within each cluster. Further details on these stimuli can be found in Appendix C (p. 447) for PAD values of words, sounds and images, and in Appendix D.2 (p. 456) for film clip URLs and cutting times.

No.	Block	Cluster	Type	Code	Description
1	1	Mildly negative	image	2410	Boy
2	1	Mildly negative	image	1240	Spider
3	1	Mildly negative	image	1230	Spider
4	1	Mildly negative	image	9210	Rain
5	1	Mildly negative	image	2206	Fingerprint
6	2	Mildly negative	word	405	solemn
7	2	Mildly negative	word	834	kick
8	2	Mildly negative	word	913	obnoxious
9	2	Mildly negative	word	806	immature
10	2	Mildly negative	word	608	skull
11	3	Mildly negative	sound	358	Writing
12	3	Mildly negative	sound	723	Radio
13	3	Mildly negative	sound	700	Toilet
14	3	Mildly negative	sound	708	Clock
15	3	Mildly negative	sound	382	Shovel
16	4	Mildly negative	film	NA	Lecture on metal solidification
17	4	Mildly negative	film	NA	Thunderstorm
18	4	Mildly negative	film	NA	Catterpillar attacked by insect
19	4	Mildly negative	film	NA	Digging and sliding through a tunnel
20	4	Mildly negative	film	NA	Student falling asleep in class
21	5	Mildly negative	film	NA	Gum stuck on shoe
22	5	Mildly negative	film	NA	Boy kick-boxing
23	5	Mildly negative	film	NA	Board meeting

24	5	Mildly negative	film	NA	Snake in a toilet
25	5	Mildly negative	film	NA	Spoiled child making demands
26	6	Mildly negative	film	NA	Radio tuning
27	6	Mildly negative	film	NA	Police make arrest
28	6	Mildly negative	film	NA	Spoiled child crying loudly
29	6	Mildly negative	film	NA	Sheep blocking mountain road
30	6	Mildly negative	film	NA	Tombstones in cemetery
31	7	Very negative	image	9187	InjuredDog
32	7	Very negative	image	9183	HurtDog[5]
33	7	Very negative	image	3110	BurnVictim
34	7	Very negative	image	9412	DeadMan
35	7	Very negative	image	3130	Mutilation
36	8	Very negative	word	616	trauma
37	8	Very negative	word	618	victim
38	8	Very negative	word	604	scared
39	8	Very negative	word	591	drown
40	8	Very negative	word	15	ambulance
41	9	Very negative	sound	115	Bees
42	9	Very negative	sound	424	CarWreck
43	9	Very negative	sound	292	MaleScream
44	9	Very negative	sound	279	Attack1
45	9	Very negative	sound	420	CarHorns
46	10	Very negative	film	NA	Aggressive snake
47	10	Very negative	film	NA	Street conflict with police
48	10	Very negative	film	NA	Bee swarm attack
49	10	Very negative	film	NA	Harbour storm and flood
50	10	Very negative	film	NA	Drainage of an abscess
51	11	Very negative	film	NA	Syrian scene after massacre
52	11	Very negative	film	NA	Pipeline explosion
53	11	Very negative	film	NA	Stressful traffic jam
54	11	Very negative	film	NA	Car crash
55	11	Very negative	film	NA	Brain surgery
56	12	Very negative	film	NA	Former school teacher evicted from his home
57	12	Very negative	film	NA	Dog attacks two girls on dust road
58	12	Very negative	film	NA	War zone shooting

[5] This IAPS image (`9183, HurtDog`) was accidentally used instead of `9414, Execution` - the latter of which had been selected based on our previous work (see Chapter 3), from page 99. Considering the average Euclidean distance between the original stimulus (i.e., `9414, Execution`) and the other stimuli in its trio, was 1.30, and that the average Euclidean distance between the stimulus accidentally replacing it (`9183 HurtDog`), and the same trio is 1.37, the difference was judged small enough for the stimulus substitution to not be a cause for concern.

59	12	Very negative	film	NA	Man nearly run over by train
60	12	Very negative	film	NA	Violent man immobilised
61	13	Positive exciting	image	8200	WaterSkier
62	13	Positive exciting	image	4668	EroticCouple
63	13	Positive exciting	image	4670	EroticCouple
64	13	Positive exciting	image	8492	Rollercoaster
65	13	Positive exciting	image	4659	EroticCouple
66	14	Positive exciting	word	152	excitement
67	14	Positive exciting	word	530	sexy
68	14	Positive exciting	word	512	erotic
69	14	Positive exciting	word	384	sex
70	14	Positive exciting	word	422	surprised
71	15	Positive exciting	sound	311	Crowd2
72	15	Positive exciting	sound	201	EroticFem1
73	15	Positive exciting	sound	360	RollerCoaster
74	15	Positive exciting	sound	815	RockNRoll
75	15	Positive exciting	sound	817	Bongos
76	16	Positive exciting	film	NA	Erotic scene: couple by the pool
77	16	Positive exciting	film	NA	Erotic scene with secretary
78	16	Positive exciting	film	NA	Erotic scene with couple in bed
79	16	Positive exciting	film	NA	Erotic scene: brunette with boyfriend
80	16	Positive exciting	film	NA	Football victory parade
81	17	Positive exciting	film	NA	Skydiving
82	17	Positive exciting	film	NA	Fireworks
83	17	Positive exciting	film	NA	Erotic scene: man kissed by woman in hotel
84	17	Positive exciting	film	NA	Surprise family reunion
85	17	Positive exciting	film	NA	Car rally
86	18	Positive exciting	film	NA	Ibiza boat party
87	18	Positive exciting	film	NA	Erotic scene: man kissed across body by woman
88	18	Positive exciting	film	NA	Erotic scene: couple near pool table
89	18	Positive exciting	film	NA	U2 Concert
90	18	Positive exciting	film	NA	Riding a roller-coaster
91	19	Positive serene	image	5200	Flowers
92	19	Positive serene	image	2314	Binoculars
93	19	Positive serene	image	2156	Family
94	19	Positive serene	image	7480	Pasta
95	19	Positive serene	image	7390	IceCream
96	20	Positive serene	word	796	humane
97	20	Positive serene	word	761	garden

98	20	Positive serene	word	786	heal
99	20	Positive serene	word	466	useful
100	20	Positive serene	word	320	politeness
101	21	Positive serene	sound	150	Seagull
102	21	Positive serene	sound	151	Robin
103	21	Positive serene	sound	809	Harp
104	21	Positive serene	sound	112	Kids1
105	21	Positive serene	sound	810	Beethoven
106	22	Positive serene	film	NA	Venice in the evening
107	22	Positive serene	film	NA	Creating artistic hand lettering on post-its
108	22	Positive serene	film	NA	Waterfalls in forest
109	22	Positive serene	film	NA	Japanese-style garden
110	22	Positive serene	film	NA	Tropical island scenes
111	23	Positive serene	film	NA	Artist playing the Earth Harp
112	23	Positive serene	film	NA	Puppy barks softly
113	23	Positive serene	film	NA	Flowers and tulips
114	23	Positive serene	film	NA	Piano on display in city square
115	23	Positive serene	film	NA	Strawberry cheesecake recipe
116	24	Positive serene	film	NA	Relaxing yoga workout
117	24	Positive serene	film	NA	Seagulls flying over beach
118	24	Positive serene	film	NA	Heart-warming family event
119	24	Positive serene	film	NA	Bird singing on a branch
120	24	Positive serene	film	NA	Angkor Wat Temple in Cambodia
121	25	Neutral	image	1908	Jellyfish
122	25	Neutral	image	1645	Wolf
123	25	Neutral	image	9422	Battleship
124	25	Neutral	image	3550.2	Coach
125	25	Neutral	image	2220	MaleFace
126	26	Neutral	word	966	rough
127	26	Neutral	word	434	theory
128	26	Neutral	word	1004	swamp
129	26	Neutral	word	957	razor
130	26	Neutral	word	23	army
131	27	Neutral	sound	722	Walking
132	27	Neutral	sound	104	Panting
133	27	Neutral	sound	130	Pig
134	27	Neutral	sound	425	Train
135	27	Neutral	sound	246	HeartBeat
136	28	Neutral	film	NA	Speedboat crossing choppy waters
137	28	Neutral	film	NA	Knife displayed

138	28	Neutral	film	NA	University basketball game
139	28	Neutral	film	NA	River with cicadas buzzing
140	28	Neutral	film	NA	Scuba diving into a lake with jellyfish
141	29	Neutral	film	NA	Dog panting on floor
142	29	Neutral	film	NA	Military training camp with obstacle run
143	29	Neutral	film	NA	Street in Athens
144	29	Neutral	film	NA	Train passing by
145	29	Neutral	film	NA	Medical assessment
146	30	Neutral	film	NA	Lecture on the temporal discounting of reward
147	30	Neutral	film	NA	Grizzly bear and then wolf crossing forest
148	30	Neutral	film	NA	Battleship open to visitors
149	30	Neutral	film	NA	Pig farm
150	30	Neutral	film	NA	Tense dialogue

4.2.3 Instruments and measures

Different instruments were used at the start vs. during the study: at the beginning of their first session, participants were asked to complete a series of questionnaires, intended to measure the extent to which their response patterns may be biased during the study. All these measures (and indeed, the entire study) were implemented in OpenSesame[6], a Python-based[7], open-source experiment builder (Mathôt, Schreij, & Theeuwes, 2012).

For instance, a set of items was included for the purpose of measuring (and later controlling) participants' usage of various types of media. These are listed below, and are accompanied by response options available to participants, as well as the associated variable names, later used in analyses:

`How often do you take photos? (variable name coding this item: PhotoFreq)`

- ☐ `Never`
- ☐ `Rarely`
- ☐ `Once a month or more`
- ☐ `Once a week or more`
- ☐ `Daily`

`How often do you watch films? (FilmFreq)`

[6] http://osdoc.cogsci.nl/

[7] Because Python uses 0-based indexing (i.e., starts counting at 0, instead of 1), the lowest score on any Likert-scales included will be 0.

☐ Never

☐ Rarely

☐ Once a month or more

☐ Once a week or more

☐ Daily

How often do you use computers? (CompFreq)

☐ Never

☐ Rarely

☐ Once a month or more

☐ Once a week or more

☐ Daily

How comfortable are you using a computer? (CompComf)

☐ Not at all comfortable

☐ (option left blank)

☐ Neither comfortable, nor uncomfortable

☐ (option left blank)

☐ Extremely comfortable

How often do you take play games on a computer/console? (GameFreq)

☐ Never

☐ Rarely

☐ Once a month or more

☐ Once a week or more

☐ Daily

How often do you take play
Massively Multiplayer Online Role-Playing Games? (MMORPGFreq)

☐ Never

☐ Rarely

☐ Once a month or more

☐ Once a week or more

☐ Daily

How often do you use online virtual worlds? (VWFreq)

☐ Never

☐ Rarely

☐ Once a month or more

☐ Once a week or more

☐ Daily

How often do you use Second Life? (SLFreq)

☐ Never

☐ Rarely

☐ Once a month or more

☐ Once a week or more

☐ Daily

Other measures were also employed at the start of the study, for similar purposes:

International Positive and Negative Affect Schedule, Short Form (I-PANAS-SF, Thompson, 2007). As a pencil-and-paper version, this questionnaire aims at measuring two constructs (general-level positive and negative affect) reliably across cultures, using just 10 items. For two similar versions of the tool - the original PANAS scale (Watson et al., 1988), and the PANAS-X (Watson & Clark, 1999) - participants performed similarly in a computerised vs. pencil-and-paper version of the tool (Brock, Barry, Lawrence, Dey, & Rolffs, 2012; Howell, Rodzon, Kurai, & Sanchez, 2010). Hence, we found it reasonable to assume the same in the case of the international version of the same scale, and implement it on a computer. Furthermore, since the sample in our study is composed of students, we expect that the major confounds in this context, i.e., computer anxiety, would be at a minimum (Weigold, Weigold, & Russell, 2013), making the computer-based application of tools less problematic.

Patient Health Questionnaire (PHQ-8, Kroenke et al., 2009). The PHQ-8 consists of 8 items checking for depressive symptoms. A previous version of this measure, the Patient Health Questionnaire (PHQ-9, Kroenke, Spitzer, & Williams, 2001) has been shown to provide valid results when implemented on a computer system (Fann et al., 2009),so it was deemed reasonable to extend this expectation to the PHQ-8 as well. In the PHQ-8, scores equal to, or over 10 can indicate depression. However in this study, our intention is to use the PHQ-8 only as a screening, rather than a diagnostic tool, in order to avoid depression-specific information processing biasing our results. Therefore, participants scoring equal to / over 10 will only be considered for exclusion, rather than automatically excluded, also depending on their scores on other measures.

Toronto Alexithymia Scale (TAS-20, Bagby, Parker, & Taylor, 1994; Bagby, Taylor, & Parker, 1994; Parker, Taylor, & Bagby, 2003; Taylor, Bagby, & Parker, 2003). The TAS-20 includes three sub-scales aimed at measuring alexithymia: "Difficulty Identifying Feelings" (DIF), "Difficulty Describing Feelings" (DDF), and "Externally-Oriented Thinking" (EOT). Of the three, only the first two appeared relevant for the current study. In addition, the third sub-scale has been shown to present reliability concerns (Kooiman, Spinhoven, & Trijsburg, 2002). This instrument (more specifically, the DIF and DDF sub-scales) were implemented in a computerised version, due to negligible differences between the pencil-and-paper and computerised versions found by Bagby, Ayearst, Morariu, Watters, and Taylor (2014). When using all three sub-scales, scores under 51 can be considered typical. However, in our case, since the full TAS-20 will not be used, no diagnostic value or cut-off score will be used with participant scores - instead, they will function as covariates.

Other pre-study measures. Also prior to testing the various stimulus modalities, we measured participants' baseline levels for Valence, Arousal and Dominance (i.e., how positive/negative, calm/alert, dominant/submissive their state was), right after entering the laboratory. These measures were collected using the Self-Assessment Manikin[8], and will be discussed shortly, as it was also used to measure emotional responses to the stimuli themselves.

In addition, *during* the experiment we administered the following tools in order to measure differences between the emotion elicitation methods:

Self-Assessment Manikin (SAM, Bradley & Lang, 1994). This is a 9-point Likert scale used to successively measure the dimensions of Valence, Arousal, and Dominance, in response to individual stimuli, regardless of modality. Thus, participants were able to rate to what extent a stimulus changed their emotional state in terms of how positive / negative, alert / relaxed, and dominant / submissive it became. These non-verbal scales are illustrated in Figure 4.1, p. 161.

AffectButton (Broekens & Brinkman, 2009, 2013). This tool represents a caricature placed within a square, which changes its facial expressions according to where participants' mouse hovers over it (or finger, for touchscreen devices) - see Figure 4.2, p. 161. Movement along its x and y axes controls Valence and Dominance, and

[8] However, due to a technical failure, for 17 of the 60 participants, this data was only logged for one of the two study sessions. Hence, we imputed the missing data using the values from the session that was indeed available for these participants. The remaining participants presented full data on these baseline measures, recorded on both study sessions.

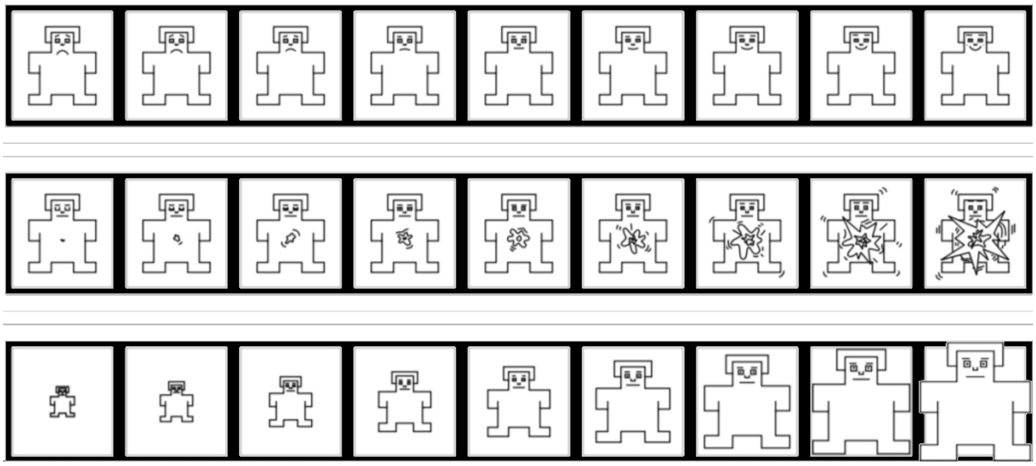


Figure 4.1: SAM scales used for rating Valence, Arousal, and Dominance levels, in response to each stimulus. Participants were asked to choose one figure from each of the three sets of images, where the first going from sad to happy, the second from bored/relaxed to alert and the third from submissive to dominant. From each set, participants have to pick one image which best describes how they would feel in a situation related to the stimulus shown.

moving closer to the square edges defines the level of Arousal. When participants identify a facial expression which they think suitably represents their current emotional state in response to a stimulus, they can submit this as their rating during the experimental task. The data saved by the tool is a set of Valence, Arousal and Dominance coordinates that control the facial expression selected, rather than the expression itself being submitted. The AffectButton was converted into an OpenSesame plugin, courtesy of Dr. Donal Stewart.

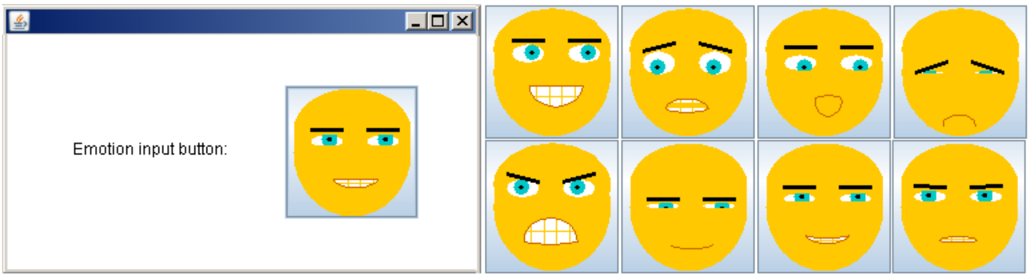


Figure 4.2: The cartoon figure above was set to change expression depending on how participants move their mouse over it. They were also instructed to click on the face to record their response, when they believed to have found an expression matching how they feel in response to a given stimulus.

Importantly, for both the SAM scales and the AffectButton, participants were asked to provide ratings for their own affective responses, which were prompted by the stimuli, rather than to evaluate the stimuli themselves - e.g., for a given stimulus, participants should rate how positive it makes them feel, rather than generally, how positively the stimulus itself should be evaluated.

ITC-Sense of Presence Inventory (ITC-SOPI, Lessiter et al., 2001). The ITC-Sense of Presence Inventory was designed to measure "presence" across a variety of media types - where presence refers to the subjective feeling of existing and operating fully within an environment (Baños et al., 2004; McMahan, 2003; Schubert et al., 2001)[9].

Four ITC-SOPI items were selected from an unpublished, short-form of the ITC-SOPI (henceforth named ITC-SOPI-SF), courtesy of Jane Lessiter (email communication on November 27, 2014)[10] - one to represent each factor identified by the authors within their original work on this instrument (Lessiter et al., 2001):

- "Sense of Physical Space": item B38, asking participants about their level of participation within the environment.
- "Engagement": item B1, asking participants how captivated they felt by the environment.
- "Ecological Validity": item B11, inquiring how credible the environment appeared to them.
- "Negative Effects": a composite item created after the publication of the ITC-SOPI, which includes multiple negative symptoms of simulator sickness, e.g., nausea, headache, dizziness, fatigue etc.

These four items were selected based on presenting the highest correlation to the scale total, being the top-loading item on their respective factor, and/or having a general meaning applicable to a variety of stimulus modalities. After each block of five images or films (but not words and sounds[11]), participants were asked to give the whole block one general rating according to these items.

4.2.4 Design and stimulus organisation

This was a within-subjects, repeated-measures design, where, by the end of the study, each participant had viewed and rated: 25 words, 25 sounds, 25 images, and 75 film clips

[9] To avoid confusion, we distinguished between "presence" and "immersion" based on the recommendation of Slater and Wilbur (1997) and Slater (1999), and referred to technological manipulations (e.g., using head-mounted displays vs. regular computer screens) as manipulations of *immersion*, rather than of *presence*, which refers solely to the subjective experience generated when using virtual environments.

[10] As per Jane Lessiter's request, these items are not reproduced here, and should be requested from the original author herself (e-mail communication, July 27, 2017), or consulted within published work (Lessiter et al., 2001).

[11] Despite our best efforts to select ITC-SOPI items which were as general in meaning as possible, these items were still not relevant to words or sounds, but rather just images and film clips.

(150 stimuli overall). The stimuli were organised into 30 stimulus blocks, with each block containing 5 items of the same modality and cluster number. To illustrate (and perhaps using an oversimplification), a participant might view a block of five successive, neutral images, and be asked to rate each of them after being displayed. This block would then be followed randomly by another block, for instance, one including five intensely negative film clips, and so on. Further details on the randomisation scheme used are under Procedure (Section 4.2.5).

Participants were invited to sign up for two study sessions, given that the total duration of the study was considered prohibitively long to be forced into a single sitting. Hence, the 30 stimulus blocks were divided randomly between the two sessions using the true random number generator at: https://www.random.org/. This resulted in the first session having been assigned a random 16 stimulus blocks (of the total number of 30), and the second session containing 14 stimulus blocks. The repartition of stimulus blocks in between sessions can be inspected in Table 4.3 (page 163).

Table 4.3: Random division of stimulus blocks (representing varying types and clusters) between the two study sessions.

Session	Type	Cluster	Number of blocks included
1	film	Mildly negative	3
1	film	Very negative	3
1	film	Positive exciting	1
1	film	Neutral	3
1	image	Very negative	1
1	image	Positive serene	1
1	sound	Positive exciting	1
1	sound	Positive serene	1
1	word	Positive exciting	1
1	word	Neutral	1
2	film	Positive exciting	2
2	film	Positive serene	3
2	image	Mildly negative	1
2	image	Positive exciting	1
2	image	Neutral	1
2	sound	Mildly negative	1
2	sound	Very negative	1
2	sound	Neutral	1
2	word	Mildly negative	1
2	word	Very negative	1
2	word	Positive serene	1

4.2.5 Procedure

The study was advertised online, via a careers / job placement university service, given that participants would receive remuneration for their time commitment (£15) at the end of the study (which lasted around 2 hours). Because the study was divided into two sessions, participants were free to choose whichever two participation slots were most suitable for them, resulting in an average gap of 1.63 days between study sessions, with a median of one day, minimum of 0 (when both sessions were done on the same day, if this was preferred), and a maximum of 10 days (SD = 2.41).

The two sessions were counterbalanced between participants, whom, on entering the laboratory, were also given an informed consent sheet, as well as a standard set of instructions to read, the content of which is reproduced in the Appendix Section D.1 (p. 454).

Once participants finished reading the instructions and gave their consent, they were able to ask the researcher for clarifications, if necessary. Subsequently, the researcher left the testing area and participants were able to begin the study, while also being told that if assistance was required, the researcher remained nearby.

The first session of the study opened with three basic questions concerning participants' age, gender and nationality. Afterwards, participants spent a few minutes answering a set of questionnaires. First, they were asked to describe their usage of various types of media, using the items mentioned previously: `PhotoFreq` (frequency of taking photos), `FilmFreq` (frequency of watching films), `CompFreq` (frequency of using a computer), `CompComf` (level of comfort when using a computer), `GameFreq` (frequency of gaming), `MMORPGFreq` (Massively Multiplayer Online Role-Playing Games gaming frequency), `VWFreq` (frequency of using virtual worlds), `SLFreq` (frequency of using Second Life).

Participants were then asked to rate their current / baseline level of affect across the PAD model, i.e., for Valence / Pleasure (how positive or negative their state is), Arousal (how relaxed / bored, or alert) and Dominance (how dominant or submissive they feel), using a computerised version of the Self-Assessment Manikin (inspired by PXLab software, Irtel, 2008; Suk, 2006; Suk & Irtel, 2010). These measures for baseline PAD levels were followed by other questionnaire items about depression (PHQ-8), general / mood-level positive and negative affect (I-PANAS-SF), and alexithymia (TAS-20).

Once answers had been provided to all these items[12], participants were able to start viewing and rating the stimulus blocks reported in Table 4.2 (page 153). These blocks

[12] Participants were not able to progress through the experiment unless complete answers were provided. This measure was implemented to eliminate the occurrence of any missing data, which could be problematic for a small-scale study, with limited funding. Nonetheless, participants were instructed that, should they wish to quit the study, they could freely do so at any time.

were presented randomly within the two study sessions. It is worth noting further that stimulus membership within a given block was fixed, however their order of presentation within their block was random, in addition to the blocks themselves being presented in random order.

This randomisation scheme (randomly displayed blocks including same-modality and same-cluster stimuli, with fixed members randomised within them at the time of presentation) was chosen for multiple reasons. Firstly, grouping several similar stimuli together within the same block reduced the risk for residual affect contaminating ratings between individual, non-congruent stimuli. If we had opted instead to randomise the stimuli themselves without nesting them within homogeneous blocks, stimuli had the potential to vary immensely from trial to trial, and create fatigue as well as spillover effects: for instance, if the image of a burn victim (code `3110` - very negative, with high Arousal) was followed by an image of ice-cream (code `7390` - very positive, with moderate Arousal), the ratings of the latter might be biased (i.e., less positive and more arousing than they might otherwise have been).

Secondly, grouping similar stimuli together also allowed us to shorten the duration of study sessions. While PAD ratings were always collected after every individual stimulus (both using the SAM scales, and the AffectButton), the four ITC-SOPI-SF items were only presented at the end of each stimulus *block*, with participants being instructed that ratings on these items should represent their holistic impression on all the stimuli included within that given block. Had ITC-SOPI-SF ratings been required for each of the 150 stimuli individually, testing times would have increased exponentially. One underlying assumption for this, however, is that the blocks are homogeneous enough to support that a single rating for a block applies equally well to the five stimuli within it.

Exposure times differed between modalities: films averaged 40s roughly, sounds lasted for 6s each, images were displayed to participants for 2s, and words, for 1s. The presentation length of films far outstripped any of the other modalities, due to film scenes taking longer to convey a situation, or for a scene to culminate. Similarly, IADS-2 sounds also were available as 6s-long audio chunks, which was sufficient for them to develop into meaningful stimuli, with the potential to influence affect. Due to the relative richness of information available, images were displayed for longer than words (2 seconds vs. 1 second). Despite the possibility of increasing the display time for all stimuli to match the films[13], this was decided against for the stimuli which did not *require* longer durations to convey meaning. This was done in order to encourage participants to give spontaneous ratings, which were less influenced by additional cognitive processing.

[13] With the exception of IADS-2 sounds, the length of which was fixed within the database.

4.3 Results

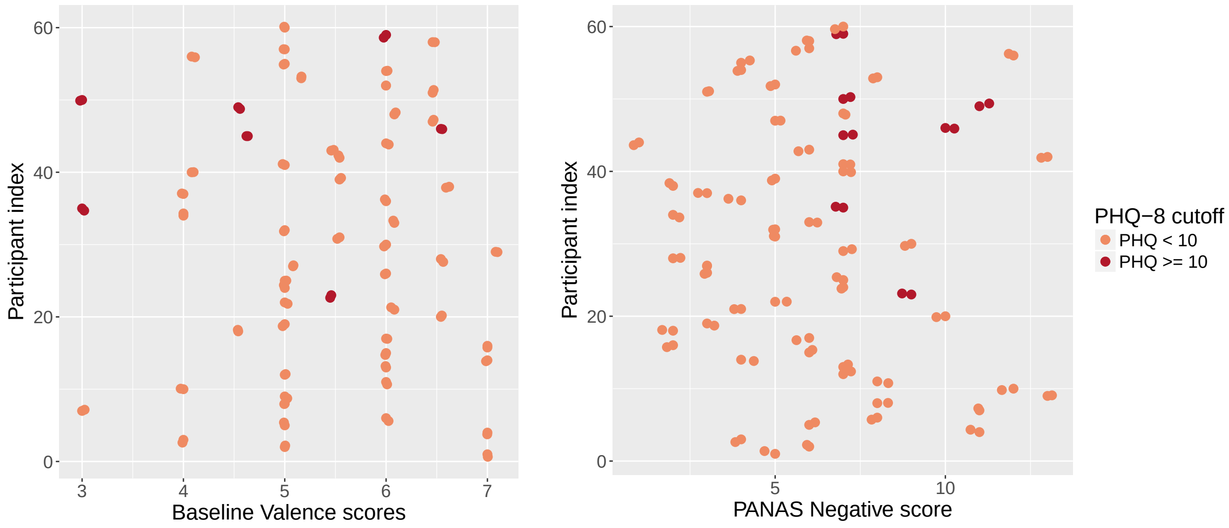
Results were computed with `R` (R Core Team, 2015), using the package versions listed in Appendix Section D.4 (p. 466).

Firstly, we decided to investigate the characteristics of our sample in terms of the measures and instruments already discussed (see Section 4.2.3). The distributions are summarised in Table 4.4, on page 166. Interestingly, seven participants scored over 10 points on the PHQ-8 depression measure, indicating a potential reason for concern. However, given that other scores (such as PANAS Negative scores and Baseline Valence) were not particularly low for these same individuals relative to the rest of the sample (see Figure 4.3), this was admitted as probable measurement error, and the data from these participants were retained.

Table 4.4: Sample description in terms of age, media usage, the PANAS schedule, PHQ, TAS-20, and baseline Valence, Arousal and Dominance, for $N = 60$.

No	Measure	Mean	Trim	Median	SD	Min	Max	Range	Skew	Kurt	SE
1	Age	23.62	22.60	23.00	5.58	18.00	46.00	28.00	1.85	3.73	0.72
2	CompComf	3.60	3.75	4.00	0.72	1.00	4.00	3.00	-1.97	3.74	0.09
3	CompFreq	3.97	4.00	4.00	0.18	3.00	4.00	1.00	-5.07	24.11	0.02
4	FilmFreq	2.60	2.58	3.00	0.76	1.00	4.00	3.00	-0.11	-0.41	0.10
5	GameFreq	1.62	1.58	2.00	1.08	0.00	4.00	4.00	0.31	-0.57	0.14
6	MMORPGFreq	0.43	0.35	0.00	0.59	0.00	2.00	2.00	0.97	-0.11	0.08
7	PhotoFreq	2.68	2.73	3.00	0.97	1.00	4.00	3.00	-0.24	-0.95	0.12
8	SLFreq	0.08	0.00	0.00	0.38	0.00	2.00	2.00	4.45	18.81	0.05
9	VWFreq	0.65	0.46	0.00	0.92	0.00	4.00	4.00	1.89	4.02	0.12
10	PANAS_Neg	6.28	6.12	6.00	2.91	1.00	13.00	12.00	0.44	-0.35	0.38
11	PANAS_Pos	13.68	13.77	14.00	2.36	7.00	18.00	11.00	-0.45	0.10	0.30
12	PHQ_total	5.15	4.65	4.00	3.77	0.00	16.00	16.00	1.14	0.88	0.49
13	TAS20_DIF	7.58	6.96	6.00	6.21	0.00	22.00	22.00	0.74	-0.43	0.80
14	TAS20_DDF	7.57	7.27	7.00	3.06	3.00	17.00	14.00	0.87	0.09	0.39
15	Base V	5.42	5.46	5.00	1.11	3.00	8.00	5.00	-0.20	-0.13	0.14
16	Base A	3.38	3.35	4.00	1.76	0.00	8.00	8.00	0.06	-0.34	0.23
17	Base D	4.02	4.06	4.00	1.23	0.00	6.00	6.00	-0.68	1.20	0.16

Note. **Base A** = Baseline Arousal; **Base D** = Baseline Dominance; **Base V** = Baseline Valence; **CompComf** = level of comfort when using a computer; **CompFreq** = frequency of using a computer; **FilmFreq** = frequency of watching films; **GameFreq** = frequency of gaming; **MMORPGFreq** = Massively Multiplayer Online Role-Playing Games gaming frequency; **PANAS_Neg** = PANAS Negative Affect scale; **PANAS_Pos** = PANAS Positive Affect scale; **PhotoFreq** = frequency of taking photos; **PHQ_total** = PHQ-8 total score; **SLFreq** = frequency of using Second Life; **TAS20_DIF** = TAS-20 Difficulty Describing Feelings subscale; **TAS20_DIF** = TAS-20 Difficulty Identifying Feeling subscale; **VWFreq** = frequency of using virtual worlds.



(a) Baseline Valence

(b) PANAS Negative Scale summed scores

(c) TAS-20 Difficulty Describing Feelings Scale

(d) TAS-20 Difficulty Identifying Feelings Scale

Figure 4.3: Participant covariate scatterplots: Valence baselines, PANAS Negative scale scores, TAS-20 DDF and DIF scales. In each scatterplot, and using a different colour, we flagged which participants received concerning PHQ-8 depression scores. Because on these measures, no obvious differences are visible between them and other participants with typical PHQ-8 scores, we attributed the difference to measurement error, and continued the analysis with the full participant pool.

4.3.1 Cultural differences

This section of the analysis only includes the 25 ANEW words, 25 IADS-2 sounds, and 25 IAPS images used, excluding any of the film clips collected from YouTube. This is because only the database stimuli are shipped with norms, and can therefore be compared to the data from our Edinburgh-based sample.

Therefore, we ran a Pearson correlation between the stimulus norms (i.e., aggregated means arising from an American student sample) made public for these stimuli, and the average PAD ratings achieved by these stimuli in our own sample. Overall across all three modalities, demonstrated strong positive relationships: for Valence ratings between the two cultures $r = 0.968$; for Arousal, $r = 0.857$; and for Dominance, $r = 0.918$. These relationships are also displayed in Figure 4.4, and are also maintained when the analysis is split by modality, in Table 4.5 (p. 169).

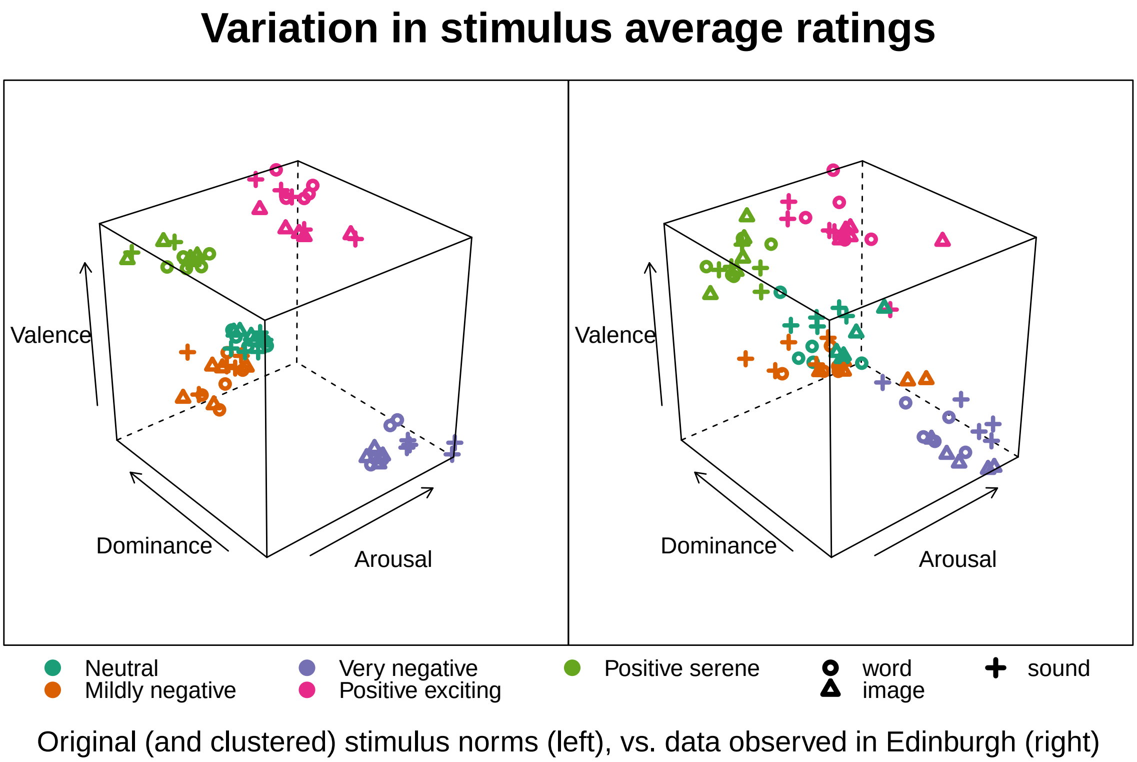


Figure 4.4: Comparison between stimulus norms and a local sample in 3D space. Despite very high cross-culture correlations between the dimensions, and the fact that the stimulus clusters tend to occupy roughly the same areas in the cube, it is immediately apparent that in the Edinburgh sample, data points are more dispersed throughout the space. Since the points displayed represent the average ratings achieved by each stimulus, it is possible that homogeneity would increase if a larger Edinburgh sample were tested.

4.3.2 Matching films to the other modalities

4.3.2.1 Pre-matching snapshot of the data

Prior to matching the 75 YouTube films to the other modalities (in order to control cross-modality variations in PAD dimensions), we visualised the PAD distributions using boxplots split by modality (including all the film clips), and included them in Figure

Table 4.5: Cross-culture correlations between PAD ratings, split by dimension and modality.

No	Dimension	Modality	r
1	Valence	word	0.97
2	Valence	image	0.98
3	Valence	sound	0.96
4	Arousal	word	0.89
5	Arousal	image	0.84
6	Arousal	sound	0.92
7	Dominance	word	0.93
8	Dominance	image	0.92
9	Dominance	sound	0.95

4.5 (page 170). ITC-SOPI-SF average values were also explored, and are displayed in Figure 4.6 (page 171) for the relevant modalities (i.e., images and films).

4.3.2.2 Stimulus matching process

In order to match 25 of the 75 film clips to the 25 available words, sounds and images, we again used the `optmatch` `R` package (version `0.9-7`, by Hansen & Klopfer, 2006), and its `fullmatch()` routine. Stimuli were matched across modalities based on the average PAD ratings for each stimulus, separately for each matching combination: words to films, sounds to films, and images to films.

Details on this procedure have already been provided previously, in Chapter 3, however in this case, we began by excluding the poorest 1/3 of matches (based on Mahalanobis distances), using two arguments available in this routine: `min.controls = 2, max.controls = 2`. This strategy had the benefit of affording the possibility of inspecting the outcome of this preliminary exclusion stage, where each treatment case (i.e., be it a word, sound, or image) was matched to the two film clips which were most similar to it in terms of their Valence, Arousal and Dominance scores. This led to a constant stratum structure of 25 groups, where each group contained: 1 treatment case (i.e., a word, sound or image), and 2 film clips (as `1:2` matches). Each time, 25 film clips would be left unmatched (as `0:1` matches).

After excluding 1/3 of the films in this way, we then pooled together all the film matches obtained across the matching combinations (i.e., the 50 film clips successfully matched to words, the 50 films matched to sounds, and another 50 matched to images - overall, 150 data points). In this pooled dataset, 9 film clips occurred only once, i.e, had been matched to only one modality of the three; 15 film clips found a match to two modalities simultaneously, but were excluded from the matching process of the third; and finally, 37 of the 75 film clips tested, found a match to each modality in the three matching operations: words, sounds, *and* images. Descriptions of these 37 film clips are

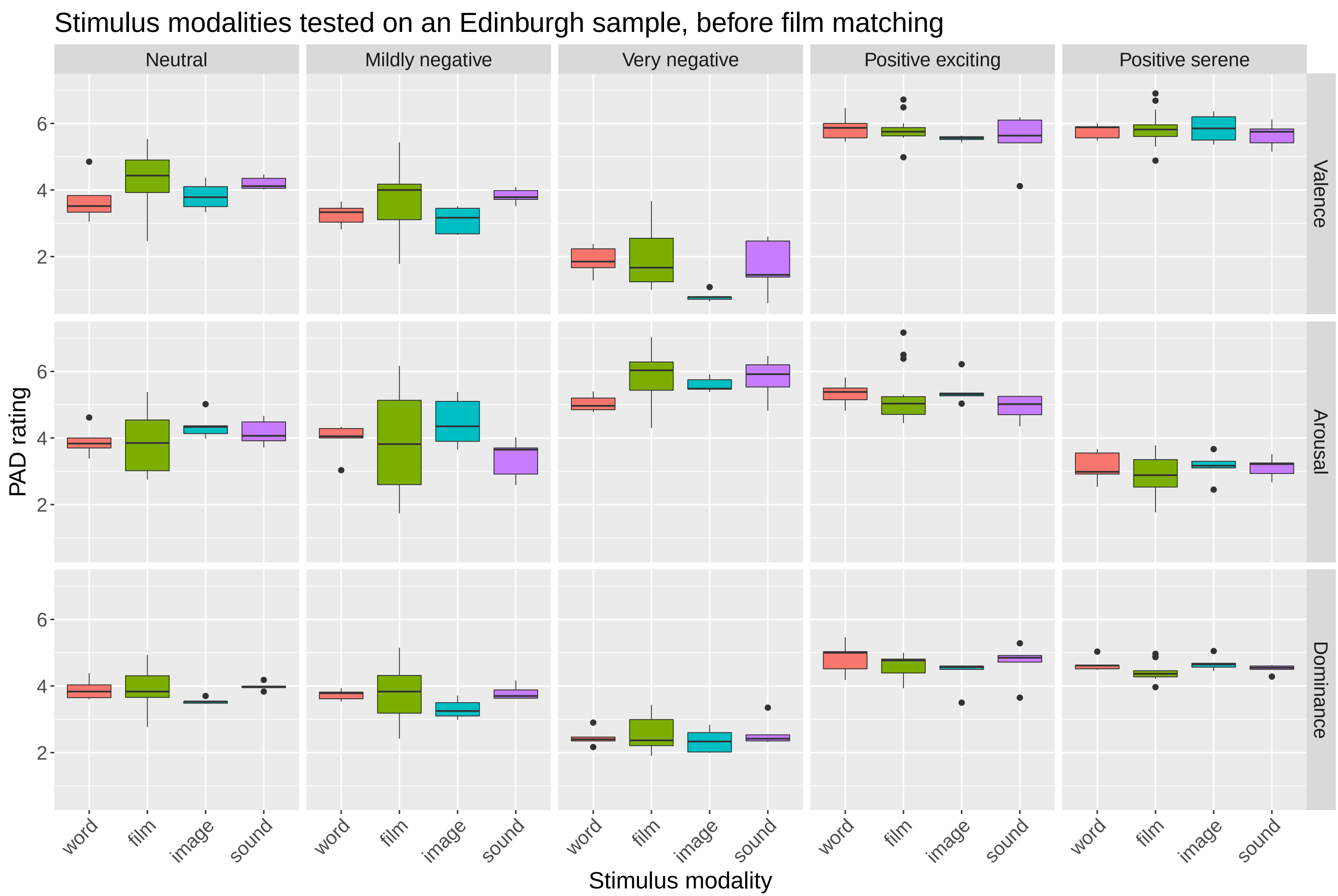


Figure 4.5: Pre-match, PAD distributions for each modality, across each PAD dimension.

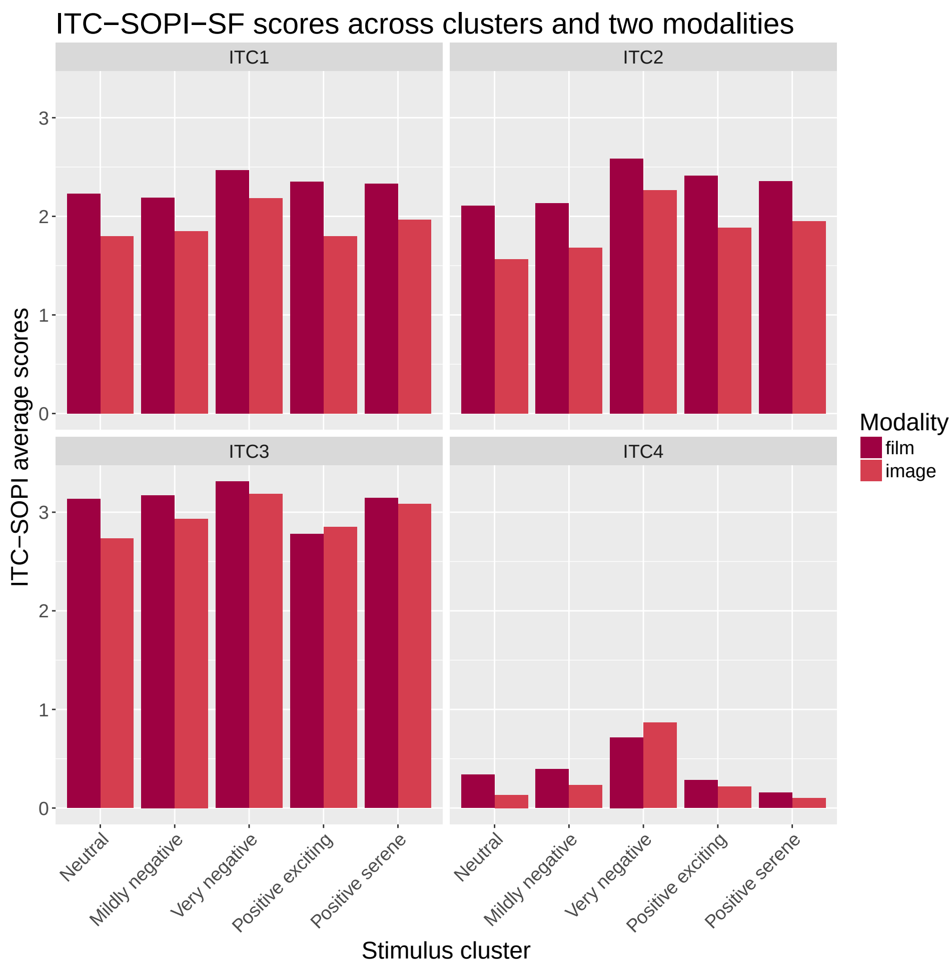


Figure 4.6: Pre-match ITC-SOPI average values for each modality where this was measured, across each item.

listed below (as well as in Appendix D.2):

```
Battleship open to visitors
Bird singing on a branch
Car crash
Caterpillar attacked by insect
Creating artistic hand lettering on post-its
Dog panting on floor
Drainage of an abscess
Erotic scene:  brunette with boyfriend
Erotic scene:  couple by the pool
Erotic scene:  couple near pool table
Erotic scene:  man kissed across body by woman
Erotic scene:  man kissed by woman in hotel
Erotic scene with couple in bed
Erotic scene with secretary
Medical assessment
Piano on display in city square
Pig farm
Puppy barks softly
Radio tuning
Relaxing yoga workout
River with cicadas buzzing
Seagulls flying over beach
Snake in a toilet
Spoiled child making demands
Strawberry cheesecake recipe
Street conflict with police
Street in Athens
Stressful traffic jam
```

```
Football victory parade                        Syrian scene after massacre
Gum stuck on shoe                              Tense dialogue
Heart-warming family event                     Train passing by
Knife displayed                                Violent man immobilised
Lecture on the temporal discounting of reward
```

Further details on these stimuli can also be found in Appendix D.3, where 12 stimuli of these 37 are shown to have been excluded, based on scrutinising their associated PAD values, and plausibility. This was considered as a defensible course of action, because the matching algorithm gives all the PAD dimensions equal weights when computing a solution. However, according to Bradley and Lang (1994), despite the theoretical importance of all the PAD factors for describing emotional experiences, they contribute differentially to explaining them, and account for 24%, 23%, and 12% of the variance in affective ratings, respectively. In the same paper, separate work by Mehrabian and Russell (1974) is discussed, and reports somewhat different percentages, of: 27%, 23%, and 14%, respectively. Hence, in the Appendix, these findings are taken into account when excluding 12 of the 37 film clips[14], and retaining just 25 - a number equal to the amount of words, sounds and images. This information is also presented visually in Figure 4.7, on page 173.

[14] Films were only considered for exclusion if their number surpassed the required $N = 5$ within each cluster.

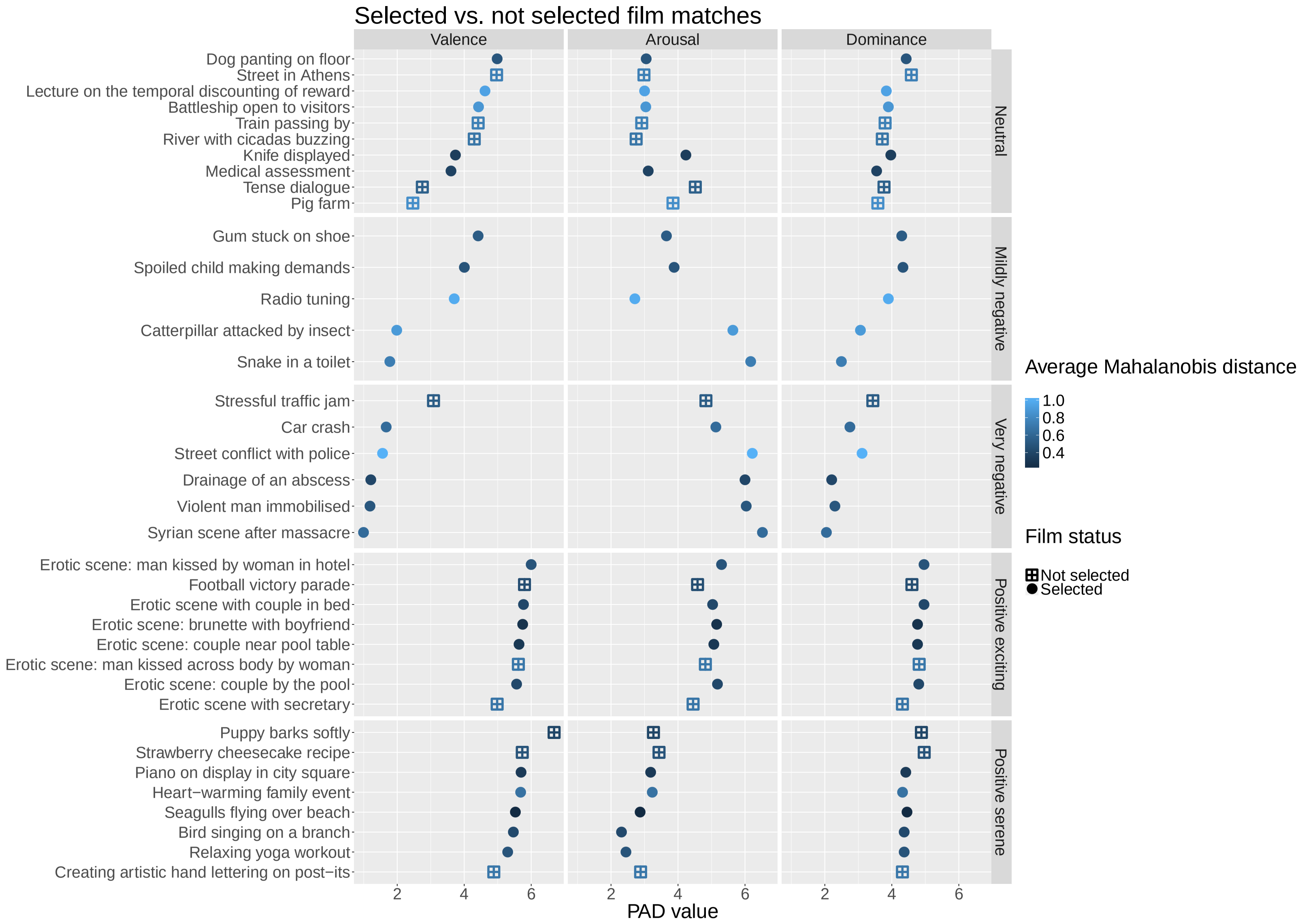


Figure 4.7: The 37 film clips (including the 25 ultimately selected) are displayed here with their *average* matching distance across the three matching operations: words to films, sounds to films, and images to films. The smaller the distance, the more suitable the match - with smaller distances coded using darker colour shades. It is worth noting that the distance measure used does not provide different weights to the three PAD dimensions, hence, on occasion (and perhaps counter-intuitively), some film clips with larger Mahalanobis distances (i.e., lighter shades) were chosen over others with a smaller distance from the other modalities. Details on the reasoning behind this appear in Appendix Section D.3 (p. 462).

4.3.2.3 Post-matching snapshot of data

Having undergone this matching procedure, the study data was reduced to contain only the 25 selected film clips, alongside the original 25 words, 25 sounds and 25 images. Recompiling two previous graphs with just this subset of the data gives Figure 4.8 (p. 177), where PAD distribution boxplots are shown, and Figure 4.9 (p. 178), where ITC-SOPI-SF scores are summarised.

Summary statistics from before and after the matching process are also presented in Table 4.7, and can be compared against those of the other stimuli (see Table 4.8).

Table 4.7: Summary statistics for films across dimensions, and split by cluster and stimulus type - before and after the matching process.

	Snapshot	Dimension	Cluster	Type	Min	Max	Mean	Median	SD
1	Pre-match	Valence	Neutral	film	2.47	5.53	4.33	4.43	0.88
2	Pre-match	Valence	Mildly negative	film	1.78	5.43	3.63	4.00	1.04
3	Pre-match	Valence	Very negative	film	1.00	3.67	1.97	1.67	0.82
4	Pre-match	Valence	Positive exciting	film	4.98	6.72	5.80	5.75	0.40
5	Pre-match	Valence	Positive serene	film	4.88	6.90	5.85	5.82	0.51
6	Pre-match	Arousal	Neutral	film	2.75	5.38	3.84	3.85	0.91
7	Pre-match	Arousal	Mildly negative	film	1.73	6.17	3.87	3.82	1.49
8	Pre-match	Arousal	Very negative	film	4.30	7.03	5.90	6.03	0.76
9	Pre-match	Arousal	Positive exciting	film	4.45	7.17	5.24	5.03	0.80
10	Pre-match	Arousal	Positive serene	film	1.77	3.78	2.93	2.88	0.57
11	Pre-match	Dominance	Neutral	film	2.77	4.93	3.95	3.83	0.55
12	Pre-match	Dominance	Mildly negative	film	2.42	5.15	3.76	3.83	0.85
13	Pre-match	Dominance	Very negative	film	1.90	3.43	2.58	2.37	0.50
14	Pre-match	Dominance	Positive exciting	film	3.93	5.00	4.61	4.77	0.33
15	Pre-match	Dominance	Positive serene	film	3.97	4.97	4.43	4.37	0.28
16	Post-match	Valence	Neutral	film	3.60	4.98	4.27	4.43	0.59
17	Post-match	Valence	Mildly negative	film	1.78	4.42	3.18	3.70	1.21
18	Post-match	Valence	Very negative	film	1.00	1.67	1.33	1.22	0.28
19	Post-match	Valence	Positive exciting	film	5.57	6.00	5.74	5.75	0.17
20	Post-match	Valence	Positive serene	film	5.30	5.70	5.54	5.53	0.17
21	Post-match	Arousal	Neutral	film	3.00	4.23	3.29	3.05	0.53
22	Post-match	Arousal	Mildly negative	film	2.72	6.17	4.41	3.88	1.44
23	Post-match	Arousal	Very negative	film	5.13	6.52	5.98	6.03	0.52
24	Post-match	Arousal	Positive exciting	film	5.03	5.30	5.15	5.15	0.11
25	Post-match	Arousal	Positive serene	film	2.32	3.23	2.81	2.87	0.42
26	Post-match	Dominance	Neutral	film	3.55	4.43	3.94	3.90	0.32
27	Post-match	Dominance	Mildly negative	film	2.50	4.33	3.62	3.90	0.81
28	Post-match	Dominance	Very negative	film	2.05	3.12	2.48	2.30	0.44
29	Post-match	Dominance	Positive exciting	film	4.77	4.97	4.85	4.80	0.10
30	Post-match	Dominance	Positive serene	film	4.32	4.45	4.38	4.37	0.05

Table 4.8: Summary statistics across dimensions, and split by cluster and stimulus type - in the case of words, sounds and images.

	Dimension	Cluster	Type	Min	Max	Mean	Median	SD
1	Valence	Neutral	word	3.05	4.85	3.72	3.52	0.69
2	Valence	Mildly negative	word	2.82	3.65	3.26	3.33	0.33
3	Valence	Very negative	word	1.28	2.38	1.88	1.85	0.44
4	Valence	Positive exciting	word	5.45	6.47	5.87	5.87	0.40
5	Valence	Positive serene	word	5.48	6.00	5.77	5.88	0.23
11	Valence	Neutral	image	3.33	4.37	3.82	3.78	0.42
12	Valence	Mildly negative	image	2.65	3.52	3.09	3.17	0.41
13	Valence	Very negative	image	0.65	1.08	0.80	0.78	0.17
14	Valence	Positive exciting	image	5.43	5.63	5.55	5.57	0.08
15	Valence	Positive serene	image	5.37	6.37	5.86	5.85	0.43
16	Valence	Neutral	sound	4.02	4.47	4.20	4.12	0.20
17	Valence	Mildly negative	sound	3.52	4.08	3.82	3.78	0.22
18	Valence	Very negative	sound	0.60	2.60	1.70	1.45	0.83
19	Valence	Positive exciting	sound	4.12	6.18	5.49	5.63	0.83
20	Valence	Positive serene	sound	5.15	6.12	5.65	5.75	0.38
21	Arousal	Neutral	word	3.38	4.62	3.91	3.83	0.46
22	Arousal	Mildly negative	word	3.03	4.33	3.94	4.05	0.53
23	Arousal	Very negative	word	4.78	5.40	5.04	4.97	0.26
24	Arousal	Positive exciting	word	4.82	5.82	5.33	5.38	0.38
25	Arousal	Positive serene	word	2.53	3.67	3.13	2.98	0.47
31	Arousal	Neutral	image	3.98	5.02	4.37	4.33	0.40
32	Arousal	Mildly negative	image	3.65	5.38	4.48	4.35	0.75
33	Arousal	Very negative	image	5.38	5.92	5.60	5.48	0.22
34	Arousal	Positive exciting	image	5.03	6.22	5.44	5.32	0.45
35	Arousal	Positive serene	image	2.45	3.67	3.14	3.17	0.44
36	Arousal	Neutral	sound	3.72	4.67	4.17	4.07	0.40
37	Arousal	Mildly negative	sound	2.58	4.02	3.37	3.65	0.60
38	Arousal	Very negative	sound	4.82	6.47	5.79	5.92	0.64
39	Arousal	Positive exciting	sound	4.35	5.27	4.92	5.02	0.39
40	Arousal	Positive serene	sound	2.67	3.52	3.12	3.22	0.33
41	Dominance	Neutral	word	3.60	4.38	3.90	3.83	0.32
42	Dominance	Mildly negative	word	3.53	3.93	3.74	3.78	0.16
43	Dominance	Very negative	word	2.17	2.90	2.46	2.40	0.27
44	Dominance	Positive exciting	word	4.18	5.47	4.84	5.00	0.50
45	Dominance	Positive serene	word	4.48	5.03	4.65	4.62	0.22
51	Dominance	Neutral	image	3.47	3.70	3.54	3.53	0.09
52	Dominance	Mildly negative	image	2.98	3.72	3.31	3.25	0.30
53	Dominance	Very negative	image	2.00	2.83	2.36	2.33	0.36
54	Dominance	Positive exciting	image	3.50	4.62	4.36	4.57	0.48
55	Dominance	Positive serene	image	4.45	5.05	4.68	4.65	0.23
56	Dominance	Neutral	sound	3.83	4.18	3.99	3.98	0.13
57	Dominance	Mildly negative	sound	3.62	4.17	3.80	3.70	0.23

58	Dominance	Very negative	sound	2.32	3.35	2.59	2.42	0.43
59	Dominance	Positive exciting	sound	3.65	5.28	4.68	4.85	0.61
60	Dominance	Positive serene	sound	4.28	4.63	4.51	4.55	0.14

4.3.3 Convergent validity of the AffectButton

If proven to be a valid measurement tool for the current research, the AffectButton may greatly simplify the rating process of PAD dimensions. In order to verify this, we computed the correlations between Valence, Arousal and Dominance distributions, when assessed the usual way with the SAM, vs. when the AffectButton was used. Results are reported in Table 4.9 (p. 176), and also shown visually in Figure 4.10 (p. 179).

Table 4.9: Pearson's r coefficients assessing the level of convergence between the SAM scales and the AffectButton, across various media types and PAD dimensions.

No	Pearson r values	Stimulus type	PAD dimension
1	0.75	Film	Valence
2	0.34	Film	Arousal
3	0.36	Film	Dominance
4	0.72	Word	Valence
5	0.27	Word	Arousal
6	0.38	Word	Dominance
7	0.74	Sound	Valence
8	0.31	Sound	Arousal
9	0.33	Sound	Dominance
10	0.74	Image	Valence
11	0.29	Image	Arousal
12	0.39	Image	Dominance
13	0.74	All	Valence
14	0.31	All	Arousal
15	0.37	All	Dominance

4.3.4 Verifying cluster membership of the stimuli, using data from on a local sample

4.3.4.1 Model-based clustering

The word, sound and image stimuli specifically used in this study were decided based on a previous model-based cluster (MBC) analysis, which, in turn, used the normative data shipped with the relevant international databases. These norms represent the average ratings that each stimulus received on the PAD model, as evaluated by an American

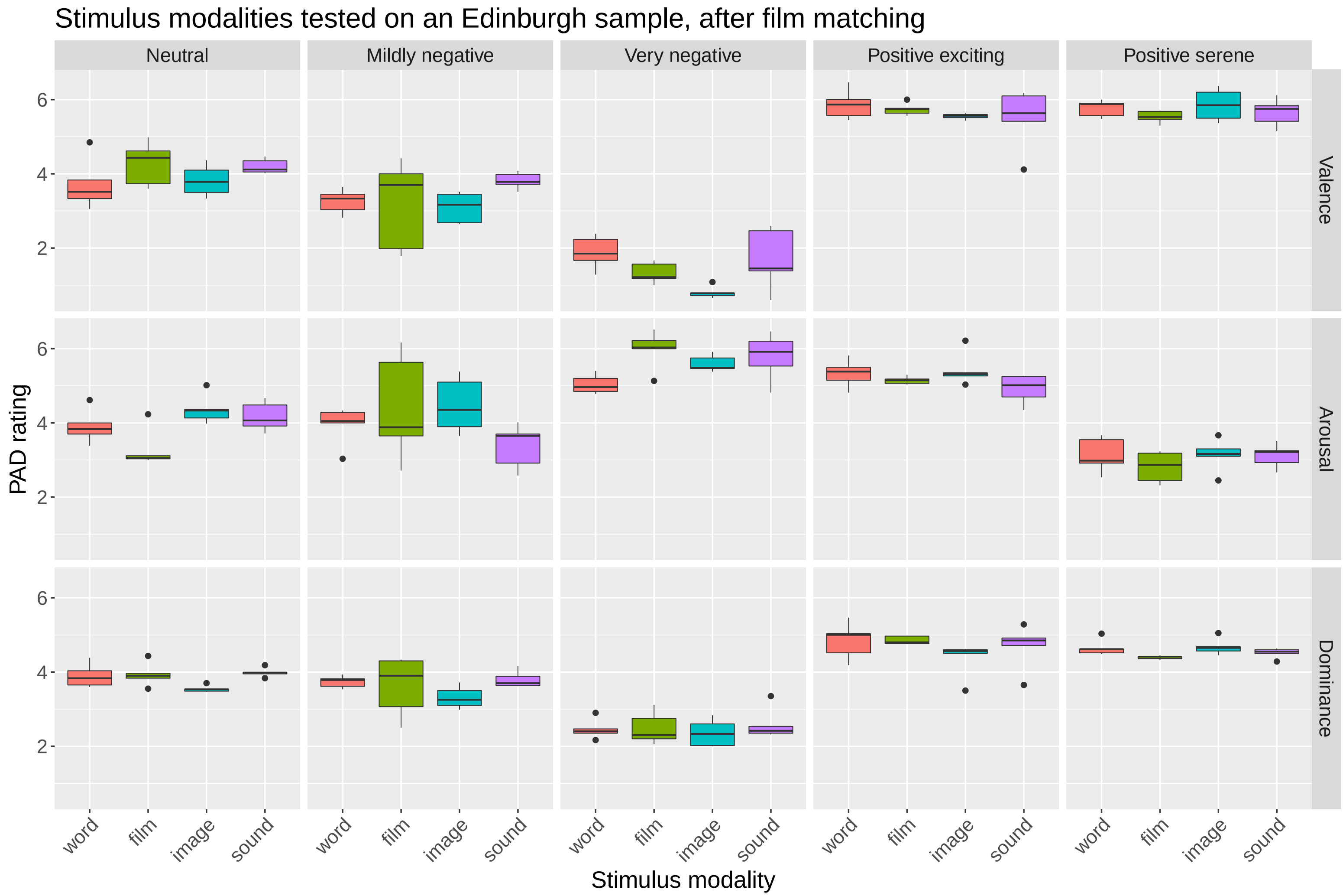


Figure 4.8: Post-match, PAD distributions for each modality, across each PAD dimension. Interestingly, there appears to be little difference in terms of boxplot alignment, relative to the previous, pre-match Figure 4.5, p. 170.

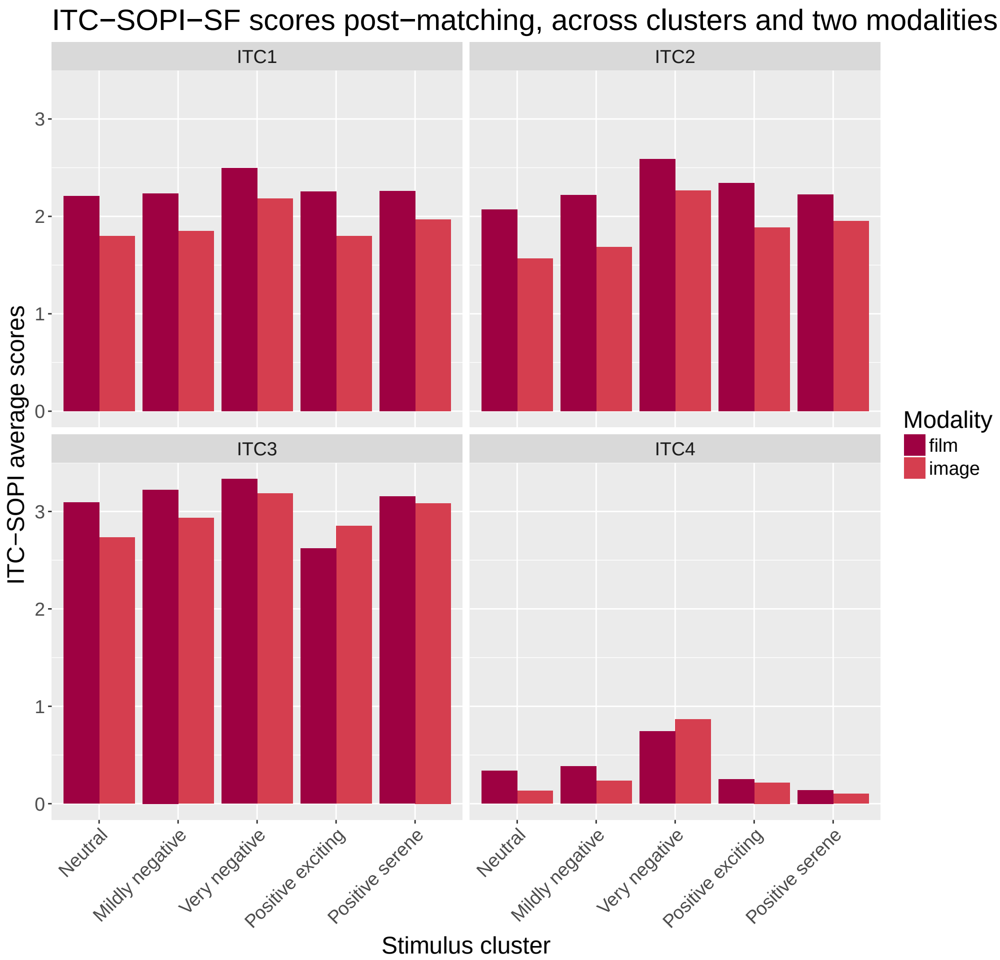


Figure 4.9: Post-match ITC-SOPI average values for each modality where this was measured, across each item. Interestingly, little difference is visible relative to the previous Figure 4.6, page 171, in that films are seen as more engaging across the board compared to images. In terms of generating physical symptoms, films were generally more associate with such effects, with the exception of only the Very negative (or Intensely negative) cluster, there images surpass films.

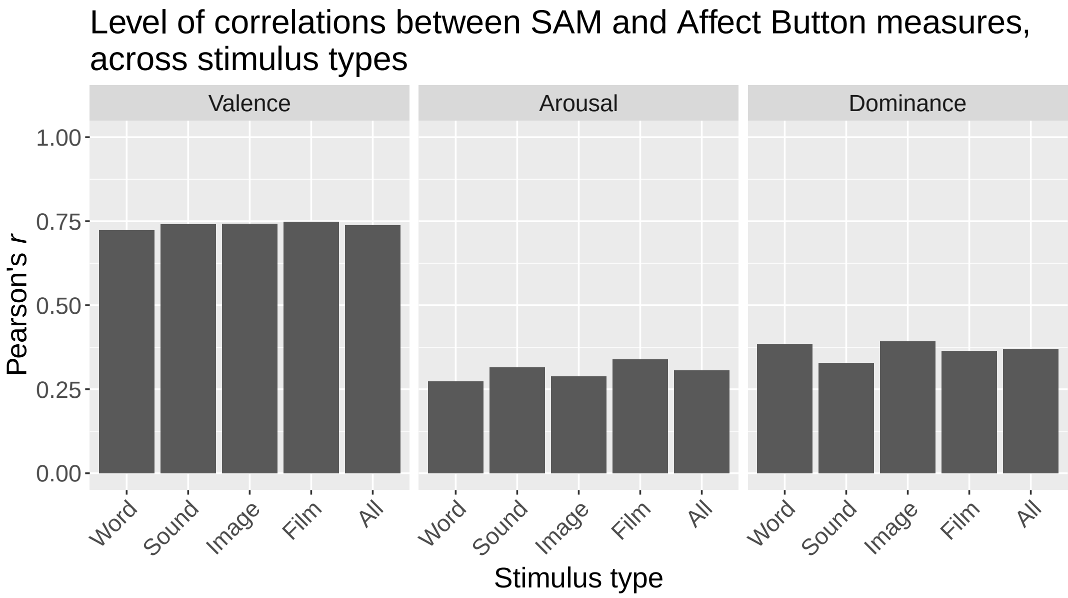


Figure 4.10: Convergent validity seems to peak in terms of Valence, which is easily recognised across both instruments. This is less the case for Arousal and Dominance.

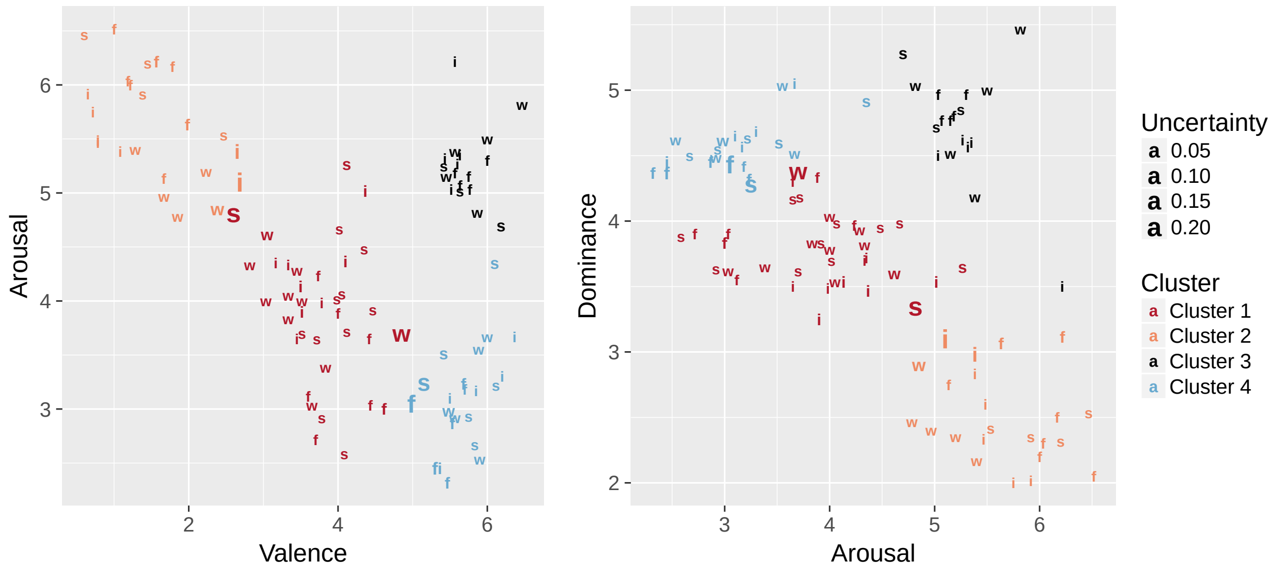
sample of approximately 100 individuals. Further details are provided in Chapter 3 (p. 99), which discusses how a structure of 4 stimulus clusters emerged from these norms (with one artificial cluster later added in manually), and also, how these 4 clusters were symmetrically arranged as: one deeply negative, one mildly negative, one positive and serene, and one positive and exciting cluster. The stimuli which best represented each of these clusters were chosen for use with our current, Edinburgh sample. Therefore, in this section, we will verify whether in a different culture, the clusters are still distinguishable as different entities.

For this purpose, we aggregated the data so that each data-point used in the MBC analysis (via `R` package `mclust`, version 5.3) represented the average SAM rating achieved by a stimulus, across all the participants. The optimal MBC model found was one where the clusters were ellipsoidal, and of equal shape (VEV), with 4 components / clusters. These clusters are described below, in Table 4.10, and illustrated in Figure 4.11 (p. 180).

Table 4.10: Model-based clustering results on data from an Edinburgh sample. Cluster centroids are presented, alongside the sample size of each cluster.

Cluster	Valence	Arousal	Dominance	N	Mixing prop
1	3.77	3.90	3.79	37	0.37
2	1.52	5.63	2.51	23	0.23
3	5.73	5.25	4.73	18	0.18
4	5.69	3.11	4.57	22	0.22

To verify this solution, we used 1000 bootstrapped samples, via the `MclustBootstrap()`



(a) Stimulus classification into the 4 clusters, within a bivariate, Valence vs. Arousal space.

(b) Stimulus classification into the 4 clusters, within a bivariate, Arousal vs. Dominance space.

Figure 4.11: Mixed-modality clusters created using model-based clustering, after including films. Despite now including films alongside the words, sounds and images, a visually similar structure emerges to our previous findings from Figure 3.11 (p. 124). Each modality is represented by its initial in the plots, and symbol size is proportional to uncertainty values.

routine in R. Results are shown graphically in Figure 4.12 (p. 181), which suggests that the bootstrapped samples follow the original solution very closely.

We also assessed the level of overlap between this solution, and the previous classification, based on the database norms. Below, Table 4.11 shows that the only major difference between the two classifications is indeed caused by the collapse of the neutral, artificial cluster, which now merged with the mildly negative one. Computing the adjusted Rand Index confirms that the similarity / overlap between classifications is substantial, with a value of 0.64.

Table 4.11: Cross-tabulation of two MBC classifications: one based on the stimulus database norms and American samples, and the other based on data collected from an Edinburgh sample.

	New cluster 1	New cluster 2	New cluster 3	New cluster 4
Neutral	19	0	0	1
Mildly negative	16	4	0	0
Very negative	1	19	0	0
Positive exciting	1	0	18	1
Positive serene	0	0	0	20

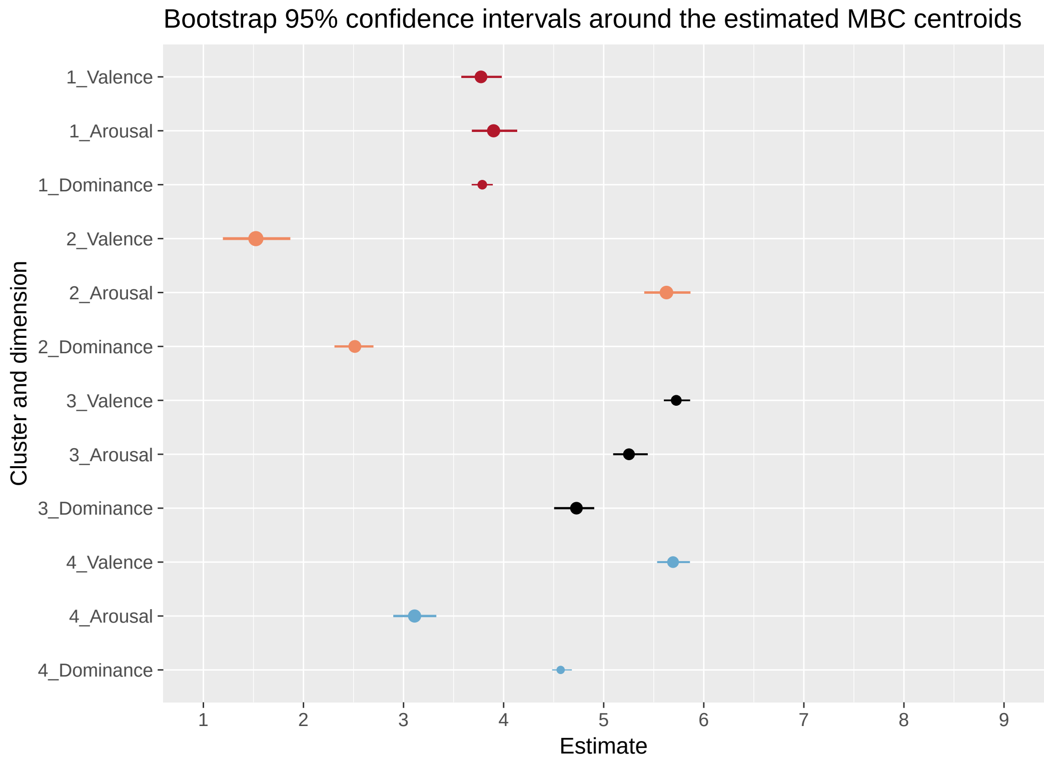


Figure 4.12: On our Edinburgh sample, the 95% bootstrapped intervals shown above tend to follow the MBC estimates (Table 4.10, p. 179) very closely. This was achieved despite the fact that, initially, the stimulus clusters were constructed using data from a foreign sample, i.e., the IAPS, IADS-2, and ANEW database norms were obtained from American students. Also, despite the fact that an extra modality - matching YouTube films - was added to these stimuli, this does not appear to have harmed the precision of the estimation, and the recovery of the same 4-cluster structure. In other words, despite substantive differences in methods and samples, the same number of 4 clusters reemerged (minus the artificial cluster, which had been manually created for use as a baseline).

4.3.4.2 Contrasts using the neutral cluster as reference

A different approach to verifying cluster structure and replicability can be achieved via generalised linear mixed models (with the `lme4::glmer()` routine in `R`). In this section, we have included an investigation where the probability of a stimulus belonging to one of two clusters, was predicted using the Valence, Arousal and Dominance ratings provided by participants, while also in the presence of covariates, i.e., any differences due to modality. We opted for this strategy of comparing our 5 clusters / categories in pairs, two-by-two, because such `glmer()` models are roughly equivalent to multinomial mixed models (where all five categories would be predicted simultaneously). In addition, such models would allow for the exclusive investigation of the contrasts of interest. Firstly, we compared the neutral cluster against every other cluster, and secondly, in a different set of models, we contrasted the two positive clusters, and the two negative clusters.

In the tables to follow, coefficients from multiple models are reported, with all models predicting the probability that a stimulus may be placed in the neutral cluster, vs. one of the following: the mildly negative cluster, the intensely negative cluster, the positive and exciting cluster, and finally, the positive and serene cluster. The predictions are based on the PAD dimensions as measured by the SAM, as well as the stimulus modality, which is expected to not influence the probability of the outcome, due to the matching process used.

A sample of the `R` code (from package `lme4`, by Bates, Maechler, Bolker, Walker, et al., 2014) used to compute these types of models is shown in Listing 4.1[15]. In terms of fixed effects, the model specification is made explicit, and shows how the binary variables are predicted from PAD ratings, as well as stimulus modality. The expectation is that stimulus modality will not have an impact on the outcome, due to the matching process carried out previously, however the PAD ratings themselves should be the main predictors for cluster membership. In terms of random effects, these were kept to a minimum in order to ensure model convergence, and include random intercepts by participant and also stimulus, as well as random slopes, where the effect of stimulus modality is allowed to vary by participant. In this last, random-slopes term, the intercept was removed in order to simplify the fitting of the model to the data, and remove the need to compute this additional parameter.

[15] Similar listings of `R` code can be found throughout this thesis. These are provided in order to facilitate the understanding / as reminders of model structure at a glance. They will also serve as brief formal clarifications for the full model specification used, before stepwise procedures filter out unnecessary (combinations of) predictors in subsequent chapters.

Listing 4.1: `R` code snippet: General form for the `glmer()` models used to predict cluster membership.

```
glmer( stimulusCluster ~ ValenceRating + ArousalRating + DominanceRating +
    stimulusModality +
        (0 + stimulusModality | participantNr) +
        (1 | participantNr) +
        (1 | stimulusCode),
      data = data, family = "binomial",
      verbose = TRUE,
      control = glmerControl( optimizer = "bobyqa",
                              optCtrl = list( maxfun = 50000 ) ) )
```

These logistic models will be duplicated, in order to assess both forms of measuring the PAD dimensions. Hence, the models will predict the dichotomies using ratings collected with either the SAM, or the AffectButton.

4.3.4.2.1 SAM outcomes. Below, in Table 4.12, the probability of a stimulus belonging to the neutral cluster relative to any other cluster, is predicted using the SAM version of the PAD ratings.

Table 4.12: Predicting cluster membership using SAM ratings: neutral contrasts.

Neutral cluster vs clusters: SAM coefficients	Mildly negative	Intensely negative	Positive exciting	Positive serene
(Intercept)	0.58	1.56	−2.50	−2.06
	(7.78)	(7.18)	(7.49)	(7.45)
Valence	−0.19	−0.66***	0.41*	0.54**
	(0.19)	(0.20)	(0.19)	(0.20)
Arousal	0.00	0.11	0.15	−0.15
	(0.13)	(0.13)	(0.13)	(0.13)
Dominance	0.02	−0.05	−0.04	0.01
	(0.17)	(0.16)	(0.18)	(0.18)
Type: film	0.05	−0.04	−0.00	−0.16
	(10.92)	(10.00)	(10.55)	(10.49)
Type: image	−0.00	−0.52	0.01	−0.01
	(10.94)	(10.00)	(10.55)	(10.45)
Type: sound	0.09	0.02	−0.01	−0.11
	(10.93)	(10.03)	(10.55)	(10.48)
AIC	133.00	130.66	132.12	131.88
BIC	242.88	240.54	242.01	241.76
Log Likelihood	-47.50	-46.33	-47.06	-46.94
Num. obs.	2400	2400	2400	2400
Num. groups: Participant	60	60	60	60
Num. groups: Stimulus code	40	40	40	40

Var: Participant × Type: word	0.00	0.00	0.00	0.00
Var: Participant × Type: film	0.00	0.00	0.00	0.00
Var: Participant × Type: image	0.00	0.00	0.00	0.00
Var: Participant × Type: sound	0.00	0.00	0.00	0.00
Cov: Participant × Type: word × Type: film	0.00	0.00	0.00	0.00
Cov: Participant × Type: word × Type: image	0.00	-0.00	0.00	0.00
Cov: Participant × Type: word × Type: sound	0.00	0.00	0.00	0.00
Cov: Participant × Type: film × Type: image	0.00	-0.00	-0.00	0.00
Cov: Participant × Type: film × Type: sound	0.00	0.00	-0.00	-0.00
Cov: Participant × Type: image × Type: sound	-0.00	0.00	-0.00	-0.00
Var: Participant (Intercept)	0.00	0.00	0.00	0.00
Var: Stimulus code (Intercept)	771.36	634.71	719.24	704.95

$^{***}p < 0.001$, $^{**}p < 0.01$, $^{*}p < 0.05$

4.3.4.2.2 AffectButton outcomes. Similarly to the previous model, the neutral-based contrasts are now predicted by the PAD dimensions, as measured using the AffectButton. Table 4.13 reports the relevant coefficients.

Table 4.13: Predicting cluster membership using AffectButton ratings: neutral contrasts.

Neutral cluster vs clusters: AffectButton coefficients	Mildly negative	Intensely negative	Positive exciting	Positive serene
(Intercept)	-0.01	-0.24	-0.20	-0.32
	(7.77)	(7.47)	(7.55)	(7.56)
Valence	-0.28	-1.04^{*}	0.78	1.15
	(0.48)	(0.50)	(0.52)	(0.60)
Arousal	0.01	0.13	0.36	0.02
	(0.35)	(0.34)	(0.35)	(0.36)
Dominance	-0.11	-0.57	0.24	0.37
	(0.46)	(0.44)	(0.44)	(0.46)
Type: film	0.02	-0.01	-0.09	-0.15
	(10.97)	(10.53)	(10.70)	(10.70)
Type: image	-0.01	-0.19	0.01	-0.13
	(10.99)	(10.56)	(10.71)	(10.68)

Type: sound	0.04	−0.04	−0.05	−0.17
	(10.98)	(10.54)	(10.71)	(10.70)
AIC	133.09	132.05	132.49	132.45
BIC	242.97	241.93	242.37	242.33
Log Likelihood	-47.54	-47.03	-47.25	-47.22
Num. obs.	2400	2400	2400	2400
Num. groups: Participant	60	60	60	60
Num. groups: Stimulus code	40	40	40	40
Var: Participant × Type: word	0.00	0.00	0.00	0.00
Var: Participant × Type: film	0.00	0.00	0.00	0.00
Var: Participant × Type: image	0.00	0.00	0.00	0.00
Var: Participant × Type: sound	0.00	0.00	0.00	0.00
Cov: Participant × Type: word × Type: film	-0.00	0.00	0.00	0.00
Cov: Participant × Type: word × Type: image	0.00	0.00	0.00	0.00
Cov: Participant × Type: word × Type: sound	-0.00	0.00	0.00	0.00
Cov: Participant × Type: film × Type: image	-0.00	-0.00	-0.00	-0.00
Cov: Participant × Type: film × Type: sound	0.00	-0.00	0.00	-0.00
Cov: Participant × Type: image × Type: sound	-0.00	-0.00	0.00	0.00
Var: Participant (Intercept)	0.00	0.00	0.00	0.00
Var: Stimulus code (Intercept)	776.54	714.91	740.91	738.21

$^{***}p < 0.001$, $^{**}p < 0.01$, $^{*}p < 0.05$

4.3.4.3 Contrasts between two forms of positive or negative material

In another set of models, we have also predicted the probability of stimuli belonging to one of two *similar* forms of Valence, i.e., on the one hand, positive and serene vs. positive and exciting, and on the other, mildly negative vs. intensely negative. We anticipate that in order to discriminate between more similar Valence categories such as these, the contribution of additional dimensions such as Arousal or Dominance might become more valuable.

Results for data measured using the SAM, and separately, using the AffectButton, are reported in Table 4.14, using the same `glmer()` fixed and random effects structure as before. Depending on which predictors were incorporated into the models, the relevant coefficients may or may not be listed within the table (e.g., for models only using data measures with the AffectButton, SAM coefficients are absent). The columns also specify what contrast was used as the outcome variable: the two forms of positive, or the two forms of negative material.

Table 4.14: Predicting cluster membership using SAM or AffectButton ratings.

SAM / AffectButton coefficients	Positive exciting vs Positive serene	Mildly negative vs Very negative	Positive exciting vs Positive serene	Mildly negative vs Very negative
(Intercept)	0.80	0.92	−0.16	−0.24
	(7.60)	(7.46)	(7.75)	(7.59)
Valence (SAM)	0.07	-0.46^{*}		
	(0.21)	(0.20)		
Arousal (SAM)	-0.31^{*}	0.10		
	(0.12)	(0.13)		
Dominance (SAM)	0.02	−0.06		
	(0.18)	(0.16)		
Type: film	−0.08	−0.27	0.03	−0.17
	(10.65)	(10.45)	(10.96)	(10.73)
Type: image	0.03	−0.43	0.04	−0.16
	(10.64)	(10.44)	(10.96)	(10.72)
Type: sound	−0.05	0.12	0.00	0.04
	(10.65)	(10.42)	(10.95)	(10.69)
Valence (AffectButton)			0.26	−0.79
			(0.68)	(0.53)
Arousal (AffectButton)			−0.31	0.16
			(0.37)	(0.35)
Dominance (AffectButton)			0.13	−0.53
			(0.45)	(0.44)
AIC	132.35	131.77	133.04	132.42
BIC	242.23	241.65	242.92	242.30
Log Likelihood	-47.18	-46.89	-47.52	-47.21
Num. obs.	2400	2400	2400	2400
Num. groups: Participant	60	60	60	60
Num. groups: Stimulus code	40	40	40	40
Var: Participant × Type: word	0.00	0.00	0.00	0.00
Var: Participant × Type: film	0.00	0.00	0.00	0.00
Var: Participant × Type: image	0.00	0.00	0.00	0.00
Var: Participant × Type: sound	0.00	0.00	0.00	0.00
Cov: Participant × Type: word × Type: film	-0.00	0.00	0.00	0.00

Cov: Participant × Type: word × Type: image	0.00	0.00	0.00	0.00
Cov: Participant × Type: word × Type: sound	0.00	0.00	0.00	0.00
Cov: Participant × Type: film × Type: image	-0.00	-0.00	0.00	-0.00
Cov: Participant × Type: film × Type: sound	0.00	0.00	-0.00	-0.00
Cov: Participant × Type: image × Type: sound	-0.00	0.00	-0.00	0.00
Var: Participant (Intercept)	0.00	0.00	0.00	0.00
Var: Stimulus code (Intercept)	732.70	698.93	773.54	736.82

$^{***}p < 0.001$, $^{**}p < 0.01$, $^{*}p < 0.05$

4.3.5 Dimensionality of the ITC-SOPI-SF

Based on a communication with the original authors of the ITC-SOPI (Lessiter et al., 2001), the expectation was that the four ITC-SOPI-SF items would each reflect a different factor/component from those identified on the full scale: "Sense of Physical Space", "Engagement", "Ecological Validity", and "Negative Effects" - particularly since the short-form items were chosen in such a way as to minimise cross-loadings. Therefore, should this expectation hold up to empirical scrutiny, it would be necessary to use the items as separate variables in further analyses. We verified this on our data by conducting a parallel analysis in R, using package `paran` and the eponymous function. With 5000 iterations specified, parallel analysis suggested that two components should be retained from the data, given they presented eigenvalues above 0 (i.e., 1.37 for the first component, and 0.13 for the second).

Based on these results, we subsequently used Principal Components Analysis (PCA) to extract the two components. Results were computed using Oblimin rotation, which affords the possibility for the components to be correlated, rather than forcing them to be orthogonal. The two components explained 65%, and 35% respectively, of the variance in the data, with low complexity (i.e., items did not cross-load too severely). The level of commonality (i.e., the variation in item scores which is explained by the components) was usually very high. These values, and the item loadings can be found in Table 4.15 (p. 188). Finally, even if the components were allowed to be correlated, their relationship still emerged as extremely weak, i.e., $r = 0.06$.

Table 4.15: Loadings of the ITC-SOPI-SF items on the two components extracted. Uniqueness is 1 - communality.

Item	Component 1	Component 2	Communality	Uniqueness	Complexity
ITC1	0.86	0.07	0.76	0.240	1.0
ITC2	0.87	0.14	0.78	0.217	1.1
ITC3	0.71	-0.37	0.62	0.383	1.5
ITC4	0.05	0.95	0.91	0.091	1.0

Overall, we took these results as an indication that the PCA solution we have computed is an adequate basis for further decisions. As such, we created a summed score to reflect the first component, based on the ITC1, ITC2 and ITC3 items which all loaded highly on this component. We labelled this summed score as general 'Engagement/Presence'. The ITC4 item which loaded onto the second component was left separate from the rest, and was labelled as a 'Negative physical symptoms' component.

4.3.6 Assessing the impact of covariates

In order to take advantage of the wealth of data recorded in this study, we explored the influence of various covariates on participant responses. The covariates available for testing were situated at three separate levels of measurement: some referred to the participants themselves (e.g., age), some to film clip characteristics (e.g., bitrate), and finally, some to word characteristics (i.e., word frequency).

Given the default listwise deletion mechanism implemented in `R`, we built separate models for these three areas of inquiry, to avoid loss of data. For example, since the bitrate measure legitimately only presents values for film clips (with values for any other modality being marked as missing), any statistical model including this measure (and using listwise deletion) will remove all cases with any missing data, prior to computing coefficients. Essentially, with the inclusion of such a variable, any results would be computed only for film clips, instead of the complete set of data. Hence, separate models were constructed for each level.

Our general approach was to first create models including only participant covariates, and based on which of these were shown to significantly influence PAD ratings, we then created film-level and word-level models, with their own specific covariates, as well as any participant-level covariates found to be of interest. This was done because participant-level covariates present complete data in this repeated-measures design, and therefore would remain unaffected even when listwise deletion is implemented (e.g., a participant's age or gender would still be present even in a subset of the data where only their affective responses to film clips are retained). Each time, these covariates were introduced into models only in the presence of variables coding the cluster of the stimuli, and their modalties - given our expectation that cluster membership should be an important predictor of PAD ratings, whereas stimulus modality should not, due to the stimulus matching process. It is worth mentioning that, while an AIC/BIC-based incremental approach to building the models was indeed considered, it proved to be unfeasible due to the sheer number of predictors. Relatedly, even an automated version of this process (via `R` package `MuMIn` by Barton, 2016, and function `dredge()`) was impractical due to the combinatorial explosion of predictor combinations, which translated into extremely long computation times (e.g., even with several days' worth of computations, `dredge()` had completed under 1% of the task).

All the covariates, as well as the stimulus cluster and modality, were used in models as fixed effects, and were standardised beforehand, due to different scales of measurement. The random effects included: random intercepts by participant, and separately, by stimulus, as well as an independently fitted component for random slopes which allow for the effect of clusters to vary by participant. In this last, random-slope term, the intercept was constrained to 0 in order to avoid computing one additional model parameter, and thus endangering convergence. Convergence is also the reason why we avoided to include the full structure of random effects, and computed models using restricted maximum likelihood (REML)[16].

The fixed effects are presented as the *model estimates* in the top half of the tables to follow, with *standard errors* printed underneath, between brackets. By dividing the two, the resulting t value can be used to determine the level of significance for coefficients[17]. AIC, BIC, and random effect variances and covariances are shown in the lower half of the tables.

Models were run while specifying a maximum number of 50,000 iterations, and `'bobyqa'` was retained as the default model optimiser. The same basic model structure was maintained when predicting Valence, Arousal, as well as Dominance ratings, across three sets of linear mixed models using the `lme4` `R` package (Bates et al., 2014), in subsections 4.3.6.1 (Participant covariates), 4.3.6.2 (Film covariates) and 4.3.6.3 (Word covariates). The formal specification of these models in `R` is made explicit in Listing 4.2.

[16] However, the disadvantage when using REML (instead of Maximum Likelihood / ML), is that parsimony measures (like the AIC) are no longer trustworthy, and cannot used to compare nested models. In our case, because all models had the same structure (i.e., nested models were not used), this was not considered to be a problem.

[17] The `R` package `lme4` does not directly compute significance levels for coefficients, however the t values which are outputted can serve as an indication for this instead.

Listing 4.2: R code snippet.

```
# General structure or lmer() models, with participant-level covariates:
s <- function(x){ scale( x, center = TRUE, scale = TRUE ) }

lmer( rateV ~ Cluster + Modality +
s(rateBaseV) + s(rateBaseA) + s(rateBaseD) +
session + s(sessionTimeDiff) +
s(Age) + Gender + Nationality +
s(PhotoFreq) + s(FilmFreq) +
s(CompComf) + s(CompFreq) + s(GameFreq) + s(MMORPGFreq) + s(VWFreq) + s(SLFreq) +
s(PANAS_Neg) + s(PANAS_Pos) + s(PHQ_total) + s(TAS20_DDF) + s(TAS20_DIF) +
s(ITC_Presence) + s(ITC_Physical) +
(0 + Cluster | Participant) + (1 | Participant) + (1 | StimulusCode),
      data = fullDat,
      verbose = TRUE,
      REML = TRUE,
      control = lmerControl(optimizer = "bobyqa",
                            optCtrl = list(maxfun = 50000) ) )
```

4.3.6.1 Participant-level covariates

Following the model structure outlined in Listing 4.2 above, two sets of complete models were constructed for predicting Valence, Arousal and Dominance scores - firstly, when these dimensions were measured with the SAM, and secondly, with the AffectButton.

4.3.6.1.1 Participant-level covariate models, with outcomes measured using the SAM. For PAD dimensions measured using the three SAM scales, the resulting coefficients are reported in Table 4.16.

Table 4.16: Participant-level covariates where the outcomes were measured with the SAM scales.

SAM Participant Coefficients	Valence	Arousal	Dominance
(Intercept)	4.05^{***}	3.42^{***}	4.04^{***}
	(0.20)	(0.34)	(0.26)
Cluster: Mildly negative	-0.73^{***}	0.11	-0.22
	(0.18)	(0.20)	(0.13)
Cluster: Very negative	-2.57^{***}	1.60^{***}	-1.38^{***}
	(0.19)	(0.23)	(0.15)
Cluster: Positive exciting	1.66^{***}	1.27^{***}	0.82^{***}
	(0.19)	(0.24)	(0.16)
Cluster: Positive serene	1.67^{***}	-0.90^{***}	0.69^{***}
	(0.19)	(0.23)	(0.15)
Modality: film	-0.05	0.03	-0.08
	(0.15)	(0.17)	(0.10)
Modality: image	-0.26	0.38^{*}	-0.25^{*}
	(0.15)	(0.17)	(0.10)
Modality: sound	0.07	-0.01	-0.01
	(0.15)	(0.17)	(0.10)
Baseline Valence	0.00	-0.06	-0.02
	(0.03)	(0.04)	(0.04)
Baseline Arousal	0.07^{**}	0.03	0.07^{*}
	(0.03)	(0.04)	(0.03)
Baseline Dominance	0.02	0.04	0.01
	(0.03)	(0.03)	(0.03)
Session 2	0.19	0.03	-0.04
	(0.11)	(0.13)	(0.08)
Time difference between sessions	-0.02	0.00	0.15
	(0.05)	(0.12)	(0.10)
Age	-0.06	-0.07	-0.11
	(0.06)	(0.14)	(0.12)
Gender: Male	0.18	-0.46	0.20
	(0.10)	(0.25)	(0.21)

Nationality: Far East	-0.08	0.15	-0.56
	(0.16)	(0.40)	(0.34)
Nationality: Middle East	-0.11	1.73	-0.26
	(0.35)	(0.89)	(0.77)
Nationality: North America	-0.13	1.40^{**}	0.16
	(0.19)	(0.48)	(0.41)
Nationality: South America	0.17	1.86	-1.12
	(0.48)	(1.20)	(1.03)
Nationality: Western Europe	-0.32^{*}	1.10^{***}	-0.13
	(0.13)	(0.33)	(0.28)
PhotoFreq	0.07	-0.17	-0.02
	(0.05)	(0.12)	(0.10)
FilmFreq	-0.02	-0.11	-0.03
	(0.05)	(0.12)	(0.10)
CompComf	0.08	0.20	-0.06
	(0.06)	(0.16)	(0.14)
CompFreq	0.03	0.20	0.02
	(0.05)	(0.12)	(0.10)
GameFreq	-0.09	-0.14	-0.01
	(0.05)	(0.13)	(0.11)
MMORPGFreq	0.04	0.03	0.14
	(0.06)	(0.15)	(0.13)
VWFreq	-0.03	-0.02	-0.02
	(0.06)	(0.14)	(0.12)
SLFreq	0.06	0.01	0.05
	(0.05)	(0.13)	(0.11)
PANAS Negative Scale	-0.15^{*}	0.59^{***}	-0.08
	(0.07)	(0.17)	(0.15)
PANAS Positive Scale	0.08	-0.10	0.10
	(0.05)	(0.13)	(0.11)
PHQ Total Score	0.03	0.17	0.25^{*}
	(0.06)	(0.14)	(0.12)
TAS20 DDF Scale	-0.06	0.30	-0.24
	(0.07)	(0.18)	(0.15)
TAS20 DIF Scale	0.16^{*}	-0.68^{***}	0.09
	(0.08)	(0.21)	(0.18)
ITC Presence/Engagement	0.06^{*}	0.21^{***}	0.05
	(0.03)	(0.03)	(0.03)
ITC Physical Symptoms	-0.05^{*}	0.05	-0.02
	(0.02)	(0.03)	(0.02)
AIC	19280.90	22095.35	20574.96
BIC	19634.86	22449.31	20928.92
Log Likelihood	-9587.45	-10994.68	-10234.48

Num. obs.	5875	5875	5875
Num. groups: Stimulus code	100	100	100
Num. groups: Participant	60	60	60
Var: Stimulus code (Intercept)	0.26	0.33	0.10
Var: Participant (Intercept)	0.00	0.33	0.15
Var: Participant × Cluster: Neutral	0.26	0.82	0.49
Var: Participant × Cluster: Mildly negative	0.28	0.40	0.27
Var: Participant × Cluster: Very negative	0.38	0.89	0.67
Var: Participant × Cluster: Positive exciting	0.46	1.06	0.92
Var: Participant × Cluster: Positive serene	0.30	1.36	1.10
Cov: Participant × Cluster: Neutral × Cluster: Mildly negative	0.21	0.47	0.25
Cov: Participant × Cluster: Neutral × Cluster: Very negative	0.11	0.32	0.29
Cov: Participant × Cluster: Neutral × Cluster: Positive exciting	0.06	0.23	0.35
Cov: Participant × Cluster: Neutral × Cluster: Positive serene	0.03	0.59	0.49
Cov: Participant × Cluster: Mildly negative × Cluster: Very negative	0.25	0.22	0.30
Cov: Participant × Cluster: Mildly negative × Cluster: Positive exciting	0.06	0.11	0.11
Cov: Participant × Cluster: Mildly negative × Cluster: Positive serene	-0.06	0.21	0.20
Cov: Participant × Cluster: Very negative × Cluster: Positive exciting	-0.01	-0.10	0.21
Cov: Participant × Cluster: Very negative × Cluster: Positive serene	-0.16	-0.40	-0.12
Cov: Participant × Cluster: Positive exciting × Cluster: Positive serene	-0.01	0.54	0.38
Var: Residual	1.34	2.14	1.68

$^{***}p < 0.001$, $^{**}p < 0.01$, $^{*}p < 0.05$

4.3.6.1.2 Participant-level covariate models, with outcomes measured using the AffectButton.

For measures collected using the AffectButton, models are reported in Table 4.17. The configuration of predictors is identical to the SAM models above.

Table 4.17: Participant-level covariates where the outcomes were measured with the Affect-Button.

AffectButton Participant Coefficients	Valence	Arousal	Dominance
(Intercept)	0.10	-0.34^{*}	-0.03
	(0.05)	(0.14)	(0.10)
Cluster: Mildly negative	-0.14^{**}	0.02	-0.09
	(0.05)	(0.06)	(0.05)
Cluster: Very negative	-0.54^{***}	0.48^{***}	-0.48^{***}
	(0.05)	(0.06)	(0.06)
Cluster: Positive exciting	0.44^{***}	0.48^{***}	0.15^{*}
	(0.05)	(0.06)	(0.06)
Cluster: Positive serene	0.43^{***}	0.19^{**}	0.25^{***}
	(0.05)	(0.06)	(0.05)
Modality: film	0.02	0.17^{**}	-0.03
	(0.04)	(0.05)	(0.05)
Modality: image	-0.02	0.17^{**}	-0.04
	(0.04)	(0.05)	(0.04)
Modality: sound	0.01	0.08	0.05
	(0.04)	(0.05)	(0.04)
Baseline Valence	-0.01	-0.05^{*}	-0.02
	(0.01)	(0.02)	(0.01)
Baseline Arousal	0.04^{***}	-0.01	0.05^{***}
	(0.01)	(0.02)	(0.01)
Baseline Dominance	-0.01	-0.03^{*}	0.04^{***}
	(0.01)	(0.01)	(0.01)
Session 2	0.00	-0.03	0.03
	(0.03)	(0.04)	(0.03)
Time difference between sessions	0.01	-0.04	0.01
	(0.02)	(0.05)	(0.03)
Age	-0.03	-0.01	-0.02
	(0.02)	(0.07)	(0.04)
Gender: Male	-0.03	0.04	0.07
	(0.03)	(0.11)	(0.07)
Nationality: Far East	-0.01	-0.00	0.04
	(0.05)	(0.18)	(0.12)
Nationality: Middle East	-0.12	0.13	0.10
	(0.12)	(0.41)	(0.26)
Nationality: North America	-0.08	0.05	0.24
	(0.06)	(0.22)	(0.14)
Nationality: South America	0.17	0.07	0.40

	(0.16)	(0.56)	(0.35)
Nationality: Western Europe	−0.06	−0.11	0.09
	(0.04)	(0.15)	(0.10)
PhotoFreq	0.02	0.09	−0.00
	(0.02)	(0.06)	(0.04)
FilmFreq	−0.02	−0.06	0.01
	(0.02)	(0.06)	(0.04)
CompComf	0.05*	0.05	0.05
	(0.02)	(0.08)	(0.05)
CompFreq	−0.02	−0.03	0.02
	(0.02)	(0.05)	(0.03)
GameFreq	−0.03	−0.06	−0.04
	(0.02)	(0.06)	(0.04)
MMORPGFreq	0.02	−0.05	−0.01
	(0.02)	(0.07)	(0.05)
VWFreq	0.01	0.02	0.00
	(0.02)	(0.07)	(0.04)
SLFreq	−0.01	0.05	−0.01
	(0.02)	(0.06)	(0.04)
PANAS Negative Scale	−0.02	−0.07	−0.07
	(0.02)	(0.08)	(0.05)
PANAS Positive Scale	−0.02	0.05	0.03
	(0.02)	(0.06)	(0.04)
PHQ Total Score	0.02	0.03	0.01
	(0.02)	(0.07)	(0.04)
TAS20 DDF Scale	−0.00	−0.01	0.06
	(0.02)	(0.08)	(0.05)
TAS20 DIF Scale	0.03	0.09	−0.04
	(0.03)	(0.10)	(0.06)
ITC Presence/Engagement	0.01	0.07***	0.01
	(0.01)	(0.01)	(0.01)
ITC Physical Symptoms	−0.02**	0.00	0.01
	(0.01)	(0.01)	(0.01)
AIC	6546.05	12073.25	8110.78
BIC	6900.01	12327.03	8464.74
Log Likelihood	-3220.02	-5998.62	-4002.39
Num. obs.	5875	5875	5875
Num. groups: Stimulus code	100	100	100
Num. groups: Participant	60	60	60
Var: Stimulus code (Intercept)	0.01	0.03	0.02
Var: Participant (Intercept)	0.00	0.12	0.01
Var: Participant × Cluster: Neutral	0.03	NA	0.08
Var: Participant × Cluster: Mildly negative	0.04	NA	0.06
Var: Participant × Cluster: Very negative	0.06	NA	0.05

Var: Participant × Cluster: Positive exciting	0.07	NA	0.10
Var: Participant × Cluster: Positive serene	0.04	NA	0.08
Cov: Participant × Cluster: Neutral × Cluster: Mildly negative	0.02	NA	0.06
Cov: Participant × Cluster: Neutral × Cluster: Very negative	0.01	NA	0.02
Cov: Participant × Cluster: Neutral × Cluster: Positive exciting	0.01	NA	0.06
Cov: Participant × Cluster: Neutral × Cluster: Positive serene	0.01	NA	0.06
Cov: Participant × Cluster: Mildly negative × Cluster: Very negative	0.02	NA	0.03
Cov: Participant × Cluster: Mildly negative × Cluster: Positive exciting	0.01	NA	0.04
Cov: Participant × Cluster: Mildly negative × Cluster: Positive serene	-0.01	NA	0.05
Cov: Participant × Cluster: Very negative × Cluster: Positive exciting	-0.02	NA	0.01
Cov: Participant × Cluster: Very negative × Cluster: Positive serene	-0.03	NA	0.01
Cov: Participant × Cluster: Positive exciting × Cluster: Positive serene	0.04	NA	0.06
Var: Residual	0.15	0.42	0.20

$^{***}p < 0.001$, $^{**}p < 0.01$, $^{*}p < 0.05$

4.3.6.2 Film-level covariates

Secondly, in terms of film-level covariates, we verified if any film-specific measures were interfering with participant ratings, but we also decided to add any subject-level predictors found to be significant previously, at any point across the models described in Tables 4.16 - 4.17 (p. 191 - 194) - particularly since in this case, it would not lead to loss of data (i.e., all cases have values on the participant-level variables). The same random-effects structure and model options were adopted as before, with coefficients reported in subsequent tables.

4.3.6.2.1 Film-level covariate models, with outcomes measured using the SAM. In Table 4.18, the models include film-level covariates, as well as any subject-level predictors previously found to be significant, with outcome variables measured using the SAM.

Table 4.18: Film-level covariates, alongside some participant measures of interest, which had a significant influence on outcomes in the previous models (Tables 4.16, and 4.17).

SAM Film coefficients	Valence	Arousal	Dominance
(Intercept)	4.39^{***}	3.13^{***}	4.01^{***}
	(0.35)	(0.47)	(0.29)
Cluster: Mildly negative	-1.09^{*}	1.08	-0.28
	(0.47)	(0.60)	(0.34)
Cluster: Very negative	-3.11^{***}	2.70^{***}	-1.52^{***}
	(0.49)	(0.64)	(0.37)
Cluster: Positive exciting	2.24^{***}	1.59	1.33^{**}
	(0.68)	(0.87)	(0.50)
Cluster: Positive serene	1.43^{**}	-0.63	0.48
	(0.48)	(0.62)	(0.35)
Film Bandwidth	0.35	-0.15	0.22
	(0.21)	(0.27)	(0.15)
Film Duration	0.08	-0.11	0.07
	(0.14)	(0.18)	(0.10)
Film FPS	0.13	-0.02	0.05
	(0.18)	(0.24)	(0.13)
Film Resolution	-0.15	0.15	-0.06
	(0.23)	(0.30)	(0.17)
Music present	-0.70	0.64	-0.36
	(0.42)	(0.54)	(0.30)
Baseline Valence	0.01	0.00	-0.04
	(0.07)	(0.10)	(0.08)
Baseline Arousal	0.12^{*}	0.31^{***}	0.05

	(0.06)	(0.09)	(0.07)
Baseline Dominance	−0.02	−0.13	0.06
	(0.06)	(0.09)	(0.07)
Nationality: Far East	−0.04	−0.41	−0.39
	(0.18)	(0.31)	(0.27)
Nationality: Middle East	−0.29	0.03	−0.02
	(0.46)	(0.78)	(0.66)
Nationality: North America	0.05	0.71	0.47
	(0.24)	(0.40)	(0.34)
Nationality: South America	−0.13	0.19	−0.65
	(0.60)	(1.01)	(0.85)
Nationality: Western Europe	−0.10	0.18	0.04
	(0.15)	(0.26)	(0.22)
CompComf	0.07	0.06	−0.05
	(0.07)	(0.12)	(0.10)
PANAS Negative Scale	−0.08	0.30^{*}	0.04
	(0.09)	(0.14)	(0.12)
PHQ Total Score	−0.01	−0.08	0.22^{*}
	(0.08)	(0.13)	(0.11)
TAS20 DIF Scale	−0.01	−0.13	-0.25^{*}
	(0.08)	(0.15)	(0.12)
ITC Presence/Engagement	0.10^{*}	0.25^{***}	0.09
	(0.05)	(0.07)	(0.06)
ITC Physical Symptoms	-0.11^{*}	0.15^{*}	−0.04
	(0.05)	(0.07)	(0.06)
AIC	5063.20	5835.82	5593.64
BIC	5286.35	6058.98	5816.79
Log Likelihood	-2489.60	-2875.91	-2754.82
Num. obs.	1500	1500	1500
Num. groups: Participant	60	60	60
Num. groups: Stimulus code	25	25	25
Var: Participant (Intercept)	0.00	0.17	0.00
Var: Participant × Cluster: Neutral	0.39	0.92	0.65
Var: Participant × Cluster: Mildly negative	0.34	0.98	0.52
Var: Participant × Cluster: Very negative	0.26	1.11	0.84
Var: Participant × Cluster: Positive exciting	1.29	1.55	1.42
Var: Participant × Cluster: Positive serene	0.38	0.83	0.58
Cov: Participant × Cluster: Neutral × Cluster: Mildly negative	0.16	0.77	0.34
Cov: Participant × Cluster: Neutral × Cluster: Very negative	0.06	0.37	0.20
Cov: Participant × Cluster: Neutral × Cluster: Positive exciting	0.05	-0.01	0.16

Cov: Participant × Cluster: Neutral × Cluster: Positive serene	0.15	0.52	0.54
Cov: Participant × Cluster: Mildly negative × Cluster: Very negative	0.20	0.77	0.51
Cov: Participant × Cluster: Mildly negative × Cluster: Positive exciting	0.11	0.19	0.08
Cov: Participant × Cluster: Mildly negative × Cluster: Positive serene	-0.02	0.32	0.12
Cov: Participant × Cluster: Very negative × Cluster: Positive exciting	-0.05	-0.06	0.29
Cov: Participant × Cluster: Very negative × Cluster: Positive serene	-0.06	-0.30	0.05
Cov: Participant × Cluster: Positive exciting × Cluster: Positive serene	0.21	-0.07	0.09
Var: Stimulus code (Intercept)	0.36	0.59	0.16
Var: Residual	1.25	2.06	1.82

$^{***}p < 0.001$, $^{**}p < 0.01$, $^{*}p < 0.05$

4.3.6.2.2 Film-level covariate models, with outcomes measured using the AffectButton.

Coefficients reported in Table 4.19 refer to outcomes measured using the AffectButton, and the previous model structure observed in the case of SAM film models, was maintained:

Table 4.19: Film-level covariates where the outcomes were measured with the AffectButton.

AffectButton Film coefficients	Valence	Arousal	Dominance
(Intercept)	0.15	-0.45^{**}	0.14
	(0.09)	(0.15)	(0.14)
Cluster: Mildly negative	-0.22	0.35^{*}	-0.26
	(0.12)	(0.18)	(0.18)
Cluster: Very negative	-0.62^{***}	0.67^{***}	-0.61^{**}
	(0.13)	(0.19)	(0.19)
Cluster: Positive exciting	0.57^{**}	0.55^{*}	0.36
	(0.17)	(0.25)	(0.26)
Cluster: Positive serene	0.40^{**}	0.28	0.25
	(0.13)	(0.18)	(0.18)
Film Bandwidth	0.07	-0.00	0.05
	(0.05)	(0.08)	(0.08)
Film Duration	0.01	-0.10	0.02
	(0.04)	(0.05)	(0.05)
Film FPS	0.00	-0.04	0.06
	(0.05)	(0.07)	(0.07)
Film Resolution	-0.01	0.02	-0.04
	(0.06)	(0.09)	(0.09)
Music present	-0.14	0.14	-0.19
	(0.11)	(0.16)	(0.16)
Baseline Valence	0.02	-0.03	-0.03
	(0.02)	(0.04)	(0.03)
Baseline Arousal	0.04^{*}	-0.02	0.09^{***}
	(0.02)	(0.03)	(0.02)
Baseline Dominance	0.02	-0.04	0.01
	(0.02)	(0.03)	(0.02)
Nationality: Far East	0.05	0.08	-0.07
	(0.06)	(0.14)	(0.09)
Nationality: Middle East	-0.02	0.30	-0.04
	(0.15)	(0.34)	(0.23)
Nationality: North America	0.01	0.11	0.22
	(0.08)	(0.18)	(0.12)
Nationality: South America	0.17	0.54	-0.07
	(0.19)	(0.44)	(0.29)
Nationality: Western Europe	-0.05	0.10	0.03

	(0.05)	(0.11)	(0.08)
CompComf	0.03	0.12^{*}	0.01
	(0.02)	(0.05)	(0.03)
PANAS Negative Scale	-0.06^{*}	-0.02	-0.01
	(0.03)	(0.06)	(0.04)
PHQ Total Score	0.03	0.05	0.01
	(0.02)	(0.06)	(0.04)
TAS20 DIF Scale	0.03	0.02	-0.03
	(0.03)	(0.06)	(0.04)
ITC Presence/Engagement	-0.00	0.12^{***}	0.01
	(0.02)	(0.03)	(0.02)
ITC Physical Symptoms	-0.07^{***}	0.02	-0.02
	(0.01)	(0.02)	(0.02)
AIC	2005.46	3125.12	2177.82
BIC	2148.92	3348.28	2400.98
Log Likelihood	-975.73	-1520.56	-1046.91
Num. obs.	1500	1500	1500
Num. groups: Participant	60	60	60
Num. groups: Stimulus code	25	25	25
Var: Participant (Intercept)	0.01	0.05	0.02
Var: Stimulus code (Intercept)	0.02	0.05	0.05
Var: Residual	0.19	0.34	0.18
Var: Participant × Cluster: Neutral	NA	0.06	0.08
Var: Participant × Cluster: Mildly negative	NA	0.08	0.03
Var: Participant × Cluster: Very negative	NA	0.14	0.03
Var: Participant × Cluster: Positive exciting	NA	0.17	0.12
Var: Participant × Cluster: Positive serene	NA	0.24	0.07
Cov: Participant × Cluster: Neutral × Cluster: Mildly negative	NA	0.03	0.04
Cov: Participant × Cluster: Neutral × Cluster: Very negative	NA	0.00	-0.00
Cov: Participant × Cluster: Neutral × Cluster: Positive exciting	NA	0.04	0.04
Cov: Participant × Cluster: Neutral × Cluster: Positive serene	NA	0.10	0.06
Cov: Participant × Cluster: Mildly negative × Cluster: Very negative	NA	0.02	0.01
Cov: Participant × Cluster: Mildly negative × Cluster: Positive exciting	NA	0.04	0.01
Cov: Participant × Cluster: Mildly negative × Cluster: Positive serene	NA	0.09	0.03
Cov: Participant × Cluster: Very negative × Cluster: Positive exciting	NA	0.07	-0.03
Cov: Participant × Cluster: Very negative × Cluster: Positive serene	NA	0.05	0.00

Cov: Participant × Cluster: Positive exciting × Cluster: Positive serene	NA	0.11	0.02

$^{***}p < 0.001$, $^{**}p < 0.01$, $^{*}p < 0.05$

4.3.6.3 Word-level covariates

Finally, the influence of covariates was also assessed at the level of the emotional words used as stimuli. Any film-specific fixed-effects were replaced with one word-specific, fixed effect (i.e., word frequency).

4.3.6.3.1 Word-level covariate models, with outcomes measured using the SAM.

The same model structure was maintained as for the previous SAM film-level models (Table 4.18, p. 197). Results are displayed in Table 4.20:

Table 4.20: Word-level covariates and their influence on outcomes measured with the SAM, alongside some participant measures of interest, which had a significant influence in the previous, participant- (Table 4.16), and film-level models (Table 4.18).

SAM Word Coefficients	Valence	Arousal	Dominance
(Intercept)	3.55^{***}	3.63^{***}	3.88^{***}
	(0.25)	(0.32)	(0.25)
Cluster: Mildly negative	-0.32	0.26	-0.02
	(0.34)	(0.33)	(0.25)
Cluster: Very negative	-1.65^{***}	1.32^{***}	-1.27^{***}
	(0.34)	(0.36)	(0.28)
Cluster: Positive exciting	2.32^{***}	1.60^{***}	1.01^{***}
	(0.31)	(0.32)	(0.24)
Cluster: Positive serene	2.18^{***}	-0.63^{*}	0.89^{***}
	(0.34)	(0.31)	(0.25)
Word Frequency	0.12	0.12	0.06
	(0.11)	(0.10)	(0.08)
Baseline Valence	-0.01	0.04	-0.10
	(0.05)	(0.11)	(0.09)
Baseline Arousal	-0.00	0.25^{**}	0.02
	(0.05)	(0.09)	(0.07)
Baseline Dominance	0.06	-0.11	0.28^{***}
	(0.05)	(0.08)	(0.07)
Nationality: Far East	0.11	-0.32	-0.69^{*}
	(0.15)	(0.36)	(0.29)
Nationality: Middle East	0.52	0.27	0.31
	(0.36)	(0.90)	(0.73)
Nationality: North America	0.06	0.70	0.22
	(0.19)	(0.47)	(0.38)
Nationality: South America	1.16	-0.83	0.92
	(0.62)	(1.24)	(1.04)
Nationality: Western Europe	0.01	0.38	0.12
	(0.12)	(0.30)	(0.25)

CompComf	0.11	0.00	−0.04
	(0.06)	(0.14)	(0.11)
PANAS Negative Scale	−0.11	0.41*	0.03
	(0.07)	(0.17)	(0.13)
PHQ Total Score	−0.01	0.30*	0.19
	(0.06)	(0.15)	(0.12)
TAS20 DIF Scale	0.11	−0.38*	−0.15
	(0.07)	(0.17)	(0.13)
ITC Presence/Engagement	0.00	0.05	0.11
	(0.04)	(0.09)	(0.07)
ITC Physical Symptoms	−0.11*	−0.06	−0.03
	(0.05)	(0.08)	(0.07)
AIC	4530.86	5243.82	4912.72
BIC	4730.82	5443.78	5112.68
Log Likelihood	-2227.43	-2583.91	-2418.36
Num. obs.	1425	1425	1425
Num. groups: Participant	60	60	60
Num. groups: Stimulus code	25	25	25
Var: Participant (Intercept)	0.00	0.49	0.33
Var: Participant × Cluster: Neutral	0.20	0.65	0.25
Var: Participant × Cluster: Mildly negative	0.65	0.82	0.24
Var: Participant × Cluster: Very negative	0.74	1.54	1.05
Var: Participant × Cluster: Positive exciting	0.64	0.84	0.68
Var: Participant × Cluster: Positive serene	0.84	0.95	0.70
Cov: Participant × Cluster: Neutral × Cluster: Mildly negative	0.22	0.23	0.07
Cov: Participant × Cluster: Neutral × Cluster: Very negative	0.14	0.01	0.05
Cov: Participant × Cluster: Neutral × Cluster: Positive exciting	0.15	-0.00	0.21
Cov: Participant × Cluster: Neutral × Cluster: Positive serene	-0.09	0.27	0.06
Cov: Participant × Cluster: Mildly negative × Cluster: Very negative	0.42	0.38	0.25
Cov: Participant × Cluster: Mildly negative × Cluster: Positive exciting	0.24	-0.09	0.08
Cov: Participant × Cluster: Mildly negative × Cluster: Positive serene	-0.12	-0.18	-0.12
Cov: Participant × Cluster: Very negative × Cluster: Positive exciting	-0.21	-0.03	-0.07
Cov: Participant × Cluster: Very negative × Cluster: Positive serene	-0.69	-0.45	-0.60
Cov: Participant × Cluster: Positive exciting × Cluster: Positive serene	0.39	-0.00	0.11

Var: Stimulus code (Intercept)	0.18	0.14	0.08
Var: Residual	1.02	1.56	1.32

$^{***}p < 0.001$, $^{**}p < 0.01$, $^{*}p < 0.05$

4.3.6.3.2 Word-level covariate models, with outcomes measured using the AffectButton.

The same word-specific models as above are reproduced here, except that outcomes were measured with the AffectButton. Coefficients are presented in Table 4.21. When predicting Dominance, however, convergence was not possible unless the term for random slopes was dropped.

Table 4.21: Word-level covariates where the outcomes were measured with the AffectButton.

AffectButton Word Coefficients	Valence	Arousal	Dominance
(Intercept)	-0.04	-0.44^{***}	0.07
	(0.06)	(0.11)	(0.09)
Cluster: Mildly negative	-0.06	0.13	-0.01
	(0.08)	(0.09)	(0.07)
Cluster: Very negative	-0.31^{***}	0.43^{***}	-0.51^{***}
	(0.08)	(0.11)	(0.07)
Cluster: Positive exciting	0.62^{***}	0.64^{***}	0.14^{*}
	(0.07)	(0.09)	(0.07)
Cluster: Positive serene	0.49^{***}	0.21^{*}	0.23^{***}
	(0.08)	(0.09)	(0.07)
Word Frequency	0.02	0.05^{*}	-0.00
	(0.02)	(0.02)	(0.02)
Baseline Valence	0.01	0.05	-0.03
	(0.02)	(0.05)	(0.03)
Baseline Arousal	0.01	-0.03	0.03
	(0.02)	(0.04)	(0.02)
Baseline Dominance	0.00	-0.09^{*}	0.05^{**}
	(0.02)	(0.04)	(0.02)
Nationality: Far East	-0.04	0.15	-0.08
	(0.05)	(0.14)	(0.12)
Nationality: Middle East	-0.05	0.14	-0.05
	(0.12)	(0.36)	(0.29)
Nationality: North America	-0.03	0.06	0.15
	(0.06)	(0.19)	(0.15)
Nationality: South America	0.19	-0.03	0.72
	(0.19)	(0.48)	(0.38)
Nationality: Western Europe	-0.01	-0.06	0.05
	(0.04)	(0.12)	(0.10)
CompComf	0.02	0.03	0.04
	(0.02)	(0.06)	(0.04)
PANAS Negative Scale	-0.03	0.02	-0.08
	(0.02)	(0.07)	(0.05)
PHQ Total Score	0.03	0.12^{*}	-0.02
	(0.02)	(0.06)	(0.05)
TAS20 DIF Scale	0.03	-0.02	-0.00

	(0.02)	(0.07)	(0.05)
ITC Presence/Engagement	0.01	0.03	0.04^{*}
	(0.02)	(0.03)	(0.02)
ITC Physical Symptoms	-0.04^{*}	-0.03	0.03
	(0.02)	(0.03)	(0.02)
AIC	1363.94	2838.89	2080.95
BIC	1563.90	3038.85	2201.97
Log Likelihood	-643.97	-1381.45	-1017.47
Num. obs.	1425	1425	1425
Num. groups: Participant	60	60	60
Num. groups: Stimulus code	25	25	25
Var: Participant (Intercept)	0.00	0.08	0.07
Var: Stimulus code (Intercept)	0.01	0.00	0.01
Var: Residual	0.11	0.30	0.21
Var: Participant × Cluster: Neutral	0.03	0.05	NA
Var: Participant × Cluster: Mildly negative	0.09	0.11	NA
Var: Participant × Cluster: Very negative	0.09	0.20	NA
Var: Participant × Cluster: Positive exciting	0.06	0.15	NA
Var: Participant × Cluster: Positive serene	0.07	0.20	NA
Cov: Participant × Cluster: Neutral × Cluster: Mildly negative	0.02	0.02	NA
Cov: Participant × Cluster: Neutral × Cluster: Very negative	0.02	-0.04	NA
Cov: Participant × Cluster: Neutral × Cluster: Positive exciting	0.01	0.01	NA
Cov: Participant × Cluster: Neutral × Cluster: Positive serene	-0.02	0.04	NA
Cov: Participant × Cluster: Mildly negative × Cluster: Very negative	0.06	0.00	NA
Cov: Participant × Cluster: Mildly negative × Cluster: Positive exciting	-0.01	-0.04	NA
Cov: Participant × Cluster: Mildly negative × Cluster: Positive serene	-0.02	0.07	NA
Cov: Participant × Cluster: Very negative × Cluster: Positive exciting	-0.04	-0.06	NA
Cov: Participant × Cluster: Very negative × Cluster: Positive serene	-0.06	0.03	NA
Cov: Participant × Cluster: Positive exciting × Cluster: Positive serene	0.03	0.02	NA

$^{***}p < 0.001$, $^{**}p < 0.01$, $^{*}p < 0.05$

4.3.6.4 Summary of covariates of interest

Based on all these tables, the covariates which at any point were able to influence **SAM** outcomes, were:

- baseline levels for Arousal and Dominance;
- participant nationality;
- scores on the negative scale of the PANAS;
- PHQ depression scores;
- the TAS-20 DIF (Difficulty Identifying Feelings) scale;
- the two ITC-SOPI-SF factor total scores (Presence/Engagement and Physical Symptoms).

In the case of **AffectButton** outcomes, the covariates of interest at any point across the tables, were:

- all three PAD baseline measures: Valence, Arousal and Dominance;
- scores on the negative scale of the PANAS;
- PHQ depression scores;
- the two ITC-SOPI-SF factor scores (Presence/Engagement and Physical Symptoms);
- user comfort when in front of a computer (item `CompComf`);
- word frequency.

4.3.6.5 Final covariate models

Based on the results computed above for each PAD dimension, and for each measurement tool (i.e., the SAM or the AffectButton), we built models containing only the significant predictors found for each scenario. On occasion, in this new configuration of effects, some previously significant predictors became non-significant, and were therefore dropped before being reported. `R` code was again used to compute these models using package `lme4` (Bates et al., 2014), using a syntax shown in Listing 4.3, and with a slightly altered (and simplified) configuration of random effects, but which was still tolerated in terms of convergence. No film covariate models are included, because no film-specific covariates were found to influence PAD ratings.

Listing 4.3: `R` code snippet: Fixed and random effects structure in the final `lmer().` models

```

lmer( rateV ~  stimulusClusterNumber + s( rateBaseA ) +
                 s(PANAS_Neg) + s(ITC_Presence_Factor) +
                     s(ITC_PhysicalSymptoms_Factor) +
                 ( stimulusClusterNumber | subject_nr ) + (1 |
                     stimulusOpenSesameCode ),
               data = fullDat, verbose = TRUE, REML = TRUE,
               control = lmerControl( optimizer = "bobyqa",
                                       optCtrl = list( maxfun = 50000 ) ) )
```

4.3.6.5.1 Participant-level models, with SAM ratings.

In the case of the SAM and participant-level covariates, Table 4.22 lists the significant predictors for each PAD dimension:

Table 4.22: Final covariate models where the outcomes were measured with the SAM. Note: AIC or BIC values for these models should not be compared directly, due to model parameters being generated through the REML method.

SAM final coefficients	Valence	Arousal	Dominance
(Intercept)	4.01^{***}	3.49^{***}	3.93^{***}
	(0.14)	(0.28)	(0.14)
Cluster: Mildly negative	-0.68^{***}	0.12	-0.22
	(0.18)	(0.20)	(0.13)
Cluster: Very negative	-2.57^{***}	1.62^{***}	-1.37^{***}
	(0.19)	(0.23)	(0.15)
Cluster: Positive exciting	1.65^{***}	1.28^{***}	0.84^{***}
	(0.20)	(0.24)	(0.16)
Cluster: Positive serene	1.67^{***}	-0.90^{***}	0.72^{***}
	(0.19)	(0.23)	(0.15)
Baseline Arousal	0.06^{*}		0.08^{*}
	(0.03)		(0.03)

ITC Presence/Engagement	0.06^{**}	0.20^{***}	
	(0.02)	(0.03)	
ITC Physical Symptoms	-0.05^{*}		
	(0.02)		
Modality: film		0.03	-0.07
		(0.17)	(0.10)
Modality: image		0.37^{*}	-0.27^{**}
		(0.17)	(0.10)
Modality: sound		-0.01	-0.00
		(0.17)	(0.10)
Nationality: Far East		-0.16	
		(0.30)	
Nationality: Middle East		1.01	
		(0.82)	
Nationality: North America		0.93^{*}	
		(0.42)	
Nationality: South America		1.31	
		(0.83)	
Nationality: Western Europe		0.70^{*}	
		(0.27)	
PANAS Negative Scale		0.62^{***}	
		(0.14)	
TAS20 DIF Scale		-0.35^{*}	
		(0.14)	
AIC	19161.99	22026.96	20897.19
BIC	19328.95	22247.35	21071.38
Log Likelihood	-9555.99	-10980.48	-10422.60
Num. obs.	5875	5875	6000
Num. groups: Stimulus code	100	100	100
Num. groups: Participant	60	60	60
Var: Stimulus code (Intercept)	0.27	0.32	0.10
Var: Participant (Intercept)	0.24	1.07	0.55
Var: Participant × Cluster: Mildly negative	0.11	0.26	0.21
Var: Participant × Cluster: Very negative	0.42	1.08	0.56
Var: Participant × Cluster: Positive exciting	0.59	1.42	0.67
Var: Participant × Cluster: Positive serene	0.50	1.00	0.63
Cov: Participant (Intercept) × Cluster: Mildly negative	-0.03	-0.25	-0.14
Cov: Participant (Intercept) × Cluster: Very negative	-0.15	-0.44	-0.07

Cov: Participant (Intercept) × Cluster: Positive exciting	-0.14	-0.57	-0.12
Cov: Participant (Intercept) × Cluster: Positive serene	-0.22	-0.22	-0.11
Cov: Participant × Cluster: Mildly negative × Cluster: Very negative	0.18	0.24	0.22
Cov: Participant × Cluster: Mildly negative × Cluster: Positive exciting	0.04	0.22	0.01
Cov: Participant × Cluster: Mildly negative × Cluster: Positive serene	-0.05	-0.04	-0.06
Cov: Participant × Cluster: Very negative × Cluster: Positive exciting	0.07	0.17	0.05
Cov: Participant × Cluster: Very negative × Cluster: Positive serene	-0.04	-0.49	-0.42
Cov: Participant × Cluster: Positive exciting ×Cluster: Positive serene	0.15	0.53	0.05
Var: Residual	1.34	2.14	1.68

$^{***}p < 0.001$, $^{**}p < 0.01$, $^{*}p < 0.05$

4.3.6.5.2 Participant-level models, with AffectButton ratings. The corresponding models for the AffectButton, are shown in Table 4.23 below:

Table 4.23: Final covariate models where the outcomes were measured with the AffectButton. Note: AIC or BIC values for these models should not be compared directly, due to model parameters being generated through the REML method.

AffectButton final coefficients	Valence	Arousal	Dominance
(Intercept)	0.06	-0.37^{***}	0.09
	(0.03)	(0.07)	(0.05)
Cluster: Mildly negative	-0.14^{**}	0.02	-0.07
	(0.04)	(0.06)	(0.05)
Cluster: Very negative	-0.54^{***}	0.48^{***}	-0.47^{***}
	(0.05)	(0.08)	(0.06)
Cluster: Positive exciting	0.44^{***}	0.48^{***}	0.16^{**}
	(0.05)	(0.07)	(0.06)
Cluster: Positive serene	0.43^{***}	0.20^{**}	0.25^{***}
	(0.05)	(0.07)	(0.06)
Baseline Arousal	0.03^{***}		0.04^{***}
	(0.01)		(0.01)
ITC Physical Symptoms	-0.02^{**}		
	(0.01)		
Modality: film		0.17^{**}	
		(0.05)	
Modality: image		0.17^{**}	
		(0.05)	
Modality: sound		0.08	
		(0.05)	
Baseline Valence		-0.04^{*}	
		(0.02)	
Baseline Dominance		-0.03^{*}	0.04^{***}
		(0.01)	(0.01)
ITC Presence/Engagement		0.07^{***}	
		(0.01)	
AIC	6359.97	11585.43	8122.23
BIC	6520.25	11772.43	8283.02
Log Likelihood	-3155.99	-5764.72	-4037.12
Num. obs.	5875	5875	6000
Num. groups: Stimulus code	100	100	100
Num. groups: Participant	60	60	60
Var: Stimulus code (Intercept)	0.01	0.03	0.02
Var: Participant (Intercept)	0.02	0.09	0.09
Var: Participant × Cluster: Mildly negative	0.03	0.02	0.02

Var: Participant × Cluster: Very negative	0.07	0.14	0.09
Var: Participant × Cluster: Positive exciting	0.07	0.09	0.07
Var: Participant × Cluster: Positive serene	0.05	0.05	0.03
Cov: Participant (Intercept) × Cluster: Mildly negative	-0.00	0.00	-0.02
Cov: Participant (Intercept) × Cluster: Very negative	-0.02	-0.04	-0.06
Cov: Participant (Intercept) × Cluster: Positive exciting	-0.01	-0.02	-0.02
Cov: Participant (Intercept) × Cluster: Positive serene	-0.01	0.03	-0.02
Cov: Participant × Cluster: Mildly negative × Cluster: Very negative	0.02	0.01	0.03
Cov: Participant × Cluster: Mildly negative × Cluster: Positive exciting	0.00	0.01	0.01
Cov: Participant × Cluster: Mildly negative × Cluster: Positive serene	-0.01	-0.00	0.00
Cov: Participant × Cluster: Very negative × Cluster: Positive exciting	-0.02	0.01	0.01
Cov: Participant × Cluster: Very negative × Cluster: Positive serene	-0.02	-0.02	0.01
Cov: Participant × Cluster: Positive exciting ×Cluster: Positive serene	0.04	0.03	0.02
Var: Residual	0.15	0.37	0.20

$^{***}p < 0.001$, $^{**}p < 0.01$, $^{*}p < 0.05$

4.3.6.5.3 Word-level Arousal models, with AffectButton ratings. Because none of the film-specific covariates were significant when predicting the PAD dimensions (as measured with either the SAM or the AffectButton), we deemed it unnecessary to create a 'final' model for films only. Similarly, the word-specific covariate we took into account - word frequency - did not affect any ratings which participants provided using the SAM. However, in the case of the AffectButton, and specifically when used to measure the Arousal dimension, word frequency did appear as significant (see Table 4.19, p. 200). Below, in Table 4.24, we report the final model for this combination of measurement tool and stimulus type:

Table 4.24: Final word-level covariate model where the Arousal ratings were measured with the AffectButton.

(Intercept)	-0.41^{***}
	(0.07)
Cluster: Mildly negative	0.13
	(0.09)
Cluster: Very negative	0.43^{***}
	(0.11)
Cluster: Positive exciting	0.60^{***}
	(0.09)
Cluster: Positive serene	0.23^{*}
	(0.09)
Word frequency	0.05^{*}
	(0.02)
PHQ Total Score	0.10^{*}
	(0.04)
AIC	2932.12
BIC	3059.64
Log Likelihood	-1442.06
Num. obs.	1500
Num. groups: Participant	60
Num. groups: Stimulus code	25
Var: Participant (Intercept)	0.13
Var: Participant × Cluster: Mildly negative	0.14
Var: Participant × Cluster: Very negative	0.36
Var: Participant × Cluster: Positive exciting	0.19
Var: Participant × Cluster: Positive serene	0.22
Cov: Participant (Intercept) × Cluster: Mildly negative	-0.03
Cov: Participant (Intercept) × Cluster: Very negative	-0.12
Cov: Participant (Intercept) × Cluster: Positive exciting	-0.04
Cov: Participant (Intercept) × Cluster: Positive serene	-0.04
Cov: Participant × Cluster: Mildly negative × Cluster: Very negative	0.11

Cov: Participant × Cluster: Mildly negative × Cluster: Positive exciting	-0.00
Cov: Participant × Cluster: Mildly negative × Cluster: Positive serene	0.09
Cov: Participant × Cluster: Very negative × Cluster: Positive exciting	0.03
Cov: Participant × Cluster: Very negative × Cluster: Positive serene	0.11
Cov: Participant × Cluster: Positive exciting × Cluster: Positive serene	0.07
Var: Stimulus code (Intercept)	0.01
Var: Residual	0.30

$^{***}p < 0.001$, $^{**}p < 0.01$, $^{*}p < 0.05$

Because it appears that only the AffectButton version of the Arousal ratings are partial to the effect of word frequency, but otherwise, measures taken with the AffectButton show fewer relationships to the rest of the recorded data, this tool was not pursued further in the current analyses - and hence, word frequency was also not considered to be a particular threat for inferences.

4.3.7 Predicting stimulus engagement

In order to investigate what contributes to a stimulus appearing more engrossing, or more likely to produce physical symptoms - especially with a view to later introduce VR as a stimulus modality for comparison - we created two further models, using package `LMERConvenienceFunctions` in `R` (Tremblay & Ransijn, 2015). Based on a full model, backward-stepwise elimination is used to filter out unnecessary predictors, in such a way as to achieve improvements in AIC values. This model-building strategy was considered reasonable for simplicity and exploratory purposes, and determined that REML could not be used in this analysis. The relevant syntax for is shown in Listing 4.4 below:

Listing 4.4: `R` code snippet backward-stepwise elimination of engagement predictors.

```r

s <- function(x){ scale( x, center = TRUE, scale = TRUE ) }

immersion_model <- lmer( ITC_Presence_Factor ~ stimulusClusterNumber + stimulusBlockType +
                           s(rateV) + s(rateA) + s(rateD) +
                           s(faceP) + s(faceA) + s(faceD) +
                           s(rateBaseV) + s(rateBaseA) + s(rateBaseD) +
                           session + s(sessionTimeDiff) +
                           s(Age) + Gender + Nationality +
                           s(PhotoFreq) + s(FilmFreq) +
                           s(CompComf) + s(CompFreq) +
                           s(GameFreq) + s(MMORPGFreq) + s(VWFreq) + s(SLFreq) +
                           s(PANAS_Neg) + s(PANAS_Pos) + s(PHQ_total) +
                           s(TAS20_DDF) + s(TAS20_DIF) +
                           (stimulusBlockType | subject_nr) +
                           (1 | stimulusOpenSesameCode),
                         data = data,
                         verbose = TRUE, REML = F,
                         control = lmerControl( optimizer = "bobyqa",
                                                optCtrl = list( maxfun = 50000 ) ) )

bfFixefLMER_F.fnc(immersion_model, method = "AIC")
```

4.3.7.1 Stimulus engagement predictors

The configuration of predictors found to be optimal in terms of AIC, during the backward-stepwise elimination, is reported in Table 4.25.

Table 4.25: Backward selection model predicting stimulus engagement.

ITC-SOPI-SF Engagement model	Coefficients
(Intercept)	5.73^{***}
	(0.39)
Cluster: Mildly negative	0.29^{*}
	(0.14)
Cluster: Very negative	1.04^{***}
	(0.16)
Cluster: Positive exciting	-0.29
	(0.15)
Cluster: Positive serene	0.51^{***}
	(0.15)
Type: film	0.90^{***}
	(0.16)
Valence (SAM)	0.19^{***}
	(0.05)
Arousal (SAM)	0.34^{***}
	(0.04)
Arousal (AffectButton)	0.20^{***}
	(0.04)
Baseline Valence	0.19^{**}
	(0.07)
Baseline Dominance	0.26^{***}
	(0.05)
Gender: Male	1.43^{**}
	(0.50)
AIC	11622.14
BIC	11724.25
Log Likelihood	-5794.07
Num. obs.	3000
Num. groups: Participant	60
Num. groups: Stimulus code	50
Var: Participant (Intercept)	4.52
Var: Participant $\times$ Type: film	1.07
Cov: Participant (Intercept) $\times$ Type: film	-0.97
Var: Stimulus code (Intercept)	0.06
Var: Residual	2.41

$^{***}p < 0.001$, $^{**}p < 0.01$, $^{*}p < 0.05$

4.3.7.2 Physical symptom predictors

Similarly, Table 4.26 shows the effects found to be useful when predicting negative physical responses to the stimuli:

Table 4.26: Backward selection model predicting Physical Symptoms, as reactions to the stimuli.

ITC-SOPI-SF Physical Symptoms model	Coefficients
(Intercept)	0.11
	(0.11)
Cluster: Mildly negative	0.07
	(0.04)
Cluster: Very negative	0.51^{***}
	(0.04)
Cluster: Positive exciting	0.04
	(0.04)
Cluster: Positive serene	-0.09^{*}
	(0.04)
Dominance (SAM)	-0.05^{***}
	(0.01)
Baseline Valence	0.10^{***}
	(0.02)
Nationality: Far East	0.55^{***}
	(0.16)
Nationality: Middle East	-0.06
	(0.44)
Nationality: North America	0.13
	(0.22)
Nationality: South America	0.01
	(0.45)
Nationality: Western Europe	-0.00
	(0.14)
PANAS Negative Scale	0.26^{***}
	(0.06)
AIC	5780.96
BIC	5877.06
Log Likelihood	-2874.48
Num. obs.	3000
Num. groups: Participant	60
Num. groups: Stimulus code	50
Var: Participant (Intercept)	0.18
Var: Stimulus code (Intercept)	0.00
Var: Residual	0.37

$^{***}p < 0.001$, $^{**}p < 0.01$, $^{*}p < 0.05$

4.4 Discussion[18]

4.4.1 Cultural differences

In this study, we were able to re-test 75 international database stimuli on a sample of students from Edinburgh. Originally, these 75 stimuli (nested within 5 clusters, each containing: 5 ANEW words, 5 IADS-2 sounds, and 5 IAPS images) had been normed on a sample of American students. In one case, this was done almost two decades ago: the ANEW word norms (or at least the associated technical report) date since 1999 (Bradley & Lang, 1999a). The IAPS (Lang et al., 2008) and IADS-2 (Bradley & Lang, 2007b) norms, on the other hand, are relatively more recent - however they too are approximately a decade old (if not more - since the time of data collection might significantly predate the year of publishing the technical reports).

Therefore, despite cultural *and* generational differences, we found that the average PAD ratings for the stimuli were highly correlated between the two samples, even when assessed separately for each modality. This is very encouraging in terms of using our stimulus selection method (see Part II, from p. 63) directly on database norms, rather than being forced to re-test stimuli locally, and then carry out the stimulus selection procedure only on the local data. A contributing factor to the high degree of convergence may also be the fact that, according to our method (e.g., Constantinescu et al., 2016), stimuli showing large SDs in the database norms were excluded before ever selecting stimuli for local use.

Finally, an interesting finding was that, while the correlations between cultures were found to be very high, the placement of average PAD ratings throughout the 3D space was more dispersed for the Edinburgh, vs. the American sample. While multiple explanations are possible, we believe this may be due to the smaller Edinburgh sample size. Had more participants been included, perhaps stimulus averages would replicate the placement of American norms to a higher extent.

4.4.2 Film matching

In order to secure an extra modality for comparing against words, sounds, and images, we collected a total number of 75 film clips, mainly from YouTube, and a few other websites. Only one third of these were ultimately matched to the other stimuli, thus leading to an overall matched set of: 25 ANEW words, 25 IADS-2 sounds, 25 IAPS images, and 25 film clips.

Before and after the filtering process through which 50 films were excluded, the

[18] Hereafter, chapter-level discussions will mostly be concerned with summarising major findings from the associated chapter. These findings will then be discussed in the context of the wider literature within the final Discussion (Chapter 9, p. 373).

PAD distributions for films changed most noticeably for Valence and Arousal, across the Neutral, Mildly negative and Very negative clusters, but remained relatively similar for the Positive and exciting, and Positive and Serene clusters, across all dimensions. This is likely due to the fact that, to begin with, the films in the neutral and two negative clusters displayed higher standard deviations across the PAD dimensions, and were more susceptible to change, and to being affected by the matching process. In contrast, the other two positive clusters were more homogeneous from the start, with the exception of a few outlying values. It was these outlying values which were usually removed/reduced by the end of the matching process.

Interestingly, values recorded for the ITC-SOPI-SF items were not affected by the matching process. The trend of films generally surpassing images in terms of perceived engagement (items ITC1, ITC2, and ITC3) was maintained post-matching across clusters with almost no exception, whereas in the case of item ITC4 (measuring negative physical reactions to the stimuli), films again tended to be rated somewhat higher, except in the case of the very negative cluster, where this trend was reverse and images were rated as more capable of eliciting such responses.

4.4.3 Verifying cluster membership

We verified on a local sample whether stimulus membership had suffered any modifications, relative to our previous analysis (see Part II). This was done in two ways: by recreating a previous cluster analysis based on stimulus norms, and by predicting a set of categorical variables (stimulus membership across various clusters) from participant PAD ratings, using generalised linear mixed-effects models.

The method of stimulus selection we originally employed was model-based cluster (MBC) analysis, on the basis of ANEW, IADS-2 and IAPS database norms. The data in this current study showed that an Edinburgh-based sample rated these stimuli very similarly to the norms. Hence, probably unsurprisingly, the 4-cluster solution developed using the stimulus norms was recreated on this dataset, with considerable overlap with the original solution (i.e., stimuli previously assigned to a cluster remained in the same cluster).

This case of replication carries particular weight, since even the addition of an extra modality - matching YouTube films - did not harm the recovery of the same 4-cluster structure, even though film clips were not part of the original analysis based on American norms, and were later simply assigned to the relevant cluster. In other words, despite substantive differences in stimulus modalities and participants, the same number of 4 clusters emerged.

The only exception to this is the artificial cluster, which had been manually created for theoretical reasons, and for serving as a baseline against which to compare the other clusters, in further analyses. Again, probably unsurprisingly, because this cluster did not emerge "naturally" as part of the original MBC solution, it was now found to have merged with the mildly negative one, i.e., its closest neighbour in 3D PAD space.

Because MBC uses different statistical algorithms relative to generalised linear mixed-effects models, some similarity, as well as some differences emerged, when comparing the results from both. When using the SAM to measure participant responses, only the Valence dimension predicted when a stimulus was more likely to be part of the neutral vs. the intensely negative / positive and exciting / positive and serene clusters. However, Valence was not able to help distinguish between the neutral and mildly negative cluster - and nor was any other dimension. As mentioned above, this is due to the two clusters being fairly close to one another in the 3D PAD space - which also led to the merger of these two in the MBC solution above.

In the same models, stimulus modality never predicted the categorical outcomes, which is to be expected due to the matching process. In addition, any random variation in results (i.e., random effects) were approximately null, with the exception of random intercepts by stimulus. Therefore, some "error" variation remains in this data, which could not be accounted for just by PAD ratings and stimulus modality.

Having investigated such "neutral-based" contrasts in these models, we then verified whether the same set of predictors could help distinguish between, on the one hand, the positive and serene and the positive and exciting clusters, and on the other hand, the mildly negative and the intensely negative cluster. Not surprisingly, we found that Arousal predicted whether a stimulus was more likely to belong to the positive and serene vs. the positive and exciting cluster, and that Valence predicted stimulus membership for the mildly vs. intensely negative cluster. Again, stimulus modality had no influence on cluster membership, and random intercepts by stimulus were non-null.

The same models as above were also run using the PAD dimensions measured with the AffectButton. With just the exception of Valence predicting the probability of a stimulus belonging to the neutral vs. intensely negative cluster, no other effects were ever significant in this case, for reasons which are not altogether clear - but will be explored in some more detail shortly.

4.4.4 AffectButton validity

Despite the AffectButton showing much promise for use in research, and particular appeal for the simple and quick measurement of all 3 PAD dimensions simultaneously, it was found to perform relatively poorly in this study, compared to the SAM.

To begin with, low correlations with the SAM scales were present particularly in the case of Arousal and Dominance, which seem to not be understood in the same way when rated on a face, vs. when rated on individual Likert scales. It is difficult to pinpoint the cause of such low convergent validity, particularly since it occurred across all stimulus types. The exception to this pattern was the Valence dimension, where the correlations between the SAM scale and the AffectButton showed much higher convergence - this pattern was also maintained irrespective of the modality in question.

Indeed, it may prove difficult to find consistency between SAM and AffectButton ratings based on facial expressions, given that: it is unlikely for all emotional states to have a corresponding facial expression, in case participants tried to find this on the AffectButton; and equally, it is unclear if all the caricature's expressions can be decoded in the same way by all participants, regardless of culture or other factors.

In addition, due to its schematic appearance, the AffectButton lacks a nose, which might have been a useful diagnostic feature for disgust- or contempt-type states. The original authors were contacted regarding this point, and indicated that participants should not be instructed to search for discrete emotions on the caricature (e.g., disgust), but rather to seek appropriate combinations of PAD dimensions. Therefore, this would render the presence of a nose unnecessary (personal communication, October 14, 2013).

Despite this reply, it is unlikely that participants can fulfil the rating task in this way, since the facial expressions would then represent a mere distraction. Furthermore, participants may not explore the responsivity space of the AffectButton fully, but rather just contend themselves with finding an expression more suitable than their last n sampled options, and then submit this as their response. No measure is set in place for determining how close these expressions were to those which participants may ideally have sought. If this is the case, this could account for the low correlations between the AffectButton and the SAM.

Finally, the PAD dimensions, when measured with the AffectButton, proved more difficult to predict in models where they were included as outcomes, or often did not function as predictors in their own right for other outcomes.

4.4.5 ITC-SOPI-SF dimensionality and predictability

Despite the original work discussed with ITC-SOPI authors involving a structure of four factors ("Sense of Physical Space", "Engagement", "Ecological Validity", and "Negative Effects"), on our Edinburgh sample this could not be replicated for the short-form items. Instead, only two factors were discovered - one labelled as general Presence/Engagement[19] with the stimuli, and the other, as Negative Physical Symptoms in response to the stimuli. This is potentially due to our sample being smaller by a factor of 10, compared to the one originally used by the authors.

Having created summed scores to reflect these two components, we explored what

[19] For this reason, when referring to ITC-SOPI-SF results, the terms of "engagement" and "presence" are used interchangeably in this work.

factors contribute to a stimulus being perceived as more engaging, or more likely to create negative physical symptoms - since both these issues will be of particular interest once Virtual Reality is introduced as a stimulus modality in the following study. Currently, however, only images and film clips were used for this analysis.

In the case of engagement, categorical cluster membership, as well as its continuous version, as the measures of Valence and Arousal (the latter of which - assessed with either the SAM or the AffectButton) contributed to the prediction: in particular, the more arousing a stimulus, the higher the level of engagement reported by participants. In the case of Valence, according to the continuous measure, the more positive a stimulus, the more engaging it seemed, however, in the case of the categorical cluster predictors, the intensely negative cluster (representing an interaction of low Valence, but high Arousal) was also significantly more engaging than the neutral one. As mentioned before based on graphic representations of the data, film clips were also generally more engaging than images. Interestingly, several other factors also contributed to the prediction, i.e., how positive or dominant the participant was feeling at baseline / when starting the study (i.e., the higher the spirits and the more confident the participant was at the start of the study, the more engaged the stimuli were perceived to be). Finally, male participants were also overall more engaged with the task than females.

In terms of negative physical symptoms, cluster membership as a categorical measure again proved useful, in that intensely negative material was more prone to eliciting physical discomfort, compared to neutral - however films were not significantly different from images in this respect. The opposite effect was registered by the positive and serene material, which was less associated with negative effects, compared to the neutral category. Interestingly, the continuous ratings for Dominance were also negatively related to the occurrence of such physical symptoms - in other words, the more confident a participant was when reacting to the stimulus, the less likely it was that they were also experiencing any negative symptoms in association with it. In addition, baseline Valence was *positively* related to this outcome; while surprising, this may be due to the fact that, on starting the study with a positive mood, any highly negative material shown would be extremely non-congruent, and might elicit an even stronger reaction than otherwise. More in line with typical expectations, however, participants scoring highly on the negative scale of the PANAS (for general negative affect) did also show stronger physical negative reactions to the stimuli. Finally, a cultural effect was also present in this model, interestingly, with Far East participants generally being more likely to exhibit (self-rated) physical symptoms, relative to Eastern Europe.

4.4.6 Influences on PAD ratings

To discover covariates that can influence PAD ratings, we opted for a simple model-building strategy due to the large amount of data available, which was hard to manage otherwise: all available predictors were entered into models separately predicting Valence, Arousal and Dominance ratings. Two versions were created for each model, depending on whether the ratings had been measured with the SAM scales or the Affect Button. Afterwards, only significant predictors were maintained into a new set of 'refined' models.

In terms of subject variables, and PAD ratings measured with the SAM, cluster membership was the only predictor to function equally across the Valence, Arousal and Dominance model. When predicting Valence scores, all clusters were found to deviate significantly from the neutral one, with the two negative clusters of course scoring lower that the neutral category, and the two positive clusters - higher. In addition, higher baseline Arousal of participants tended to increase Valence scores. Interestingly, both ITC components also served as predictors in this model, where higher engagement led to an increase in Valence ratings, and the occurrence of negative physical symptoms was unsurprisingly, associated with lower Valence. In terms of interpretation, the version of this model where the Valence dimension was measured with the AffectButton, is identical - with one exception, namely that the level of engagement with the stimuli did not influence the evaluation of Valence.

When predicting Arousal scores measured with the SAM, cluster membership was again a valuable predictor, except that the mildly negative cluster was no longer found to differ from the neutral one. Baseline Arousal now failed to contribute to the prediction, similarly to the ITC Physical Symptoms Component. However, as for Valence, engagement was still associated with higher Arousal ratings. A new set of predictors now appear to be of interest, with images now seen as more arousing than words (used as a baseline category); the other two modalities, i.e., films and sounds, did not differ from the affective words in terms of Arousal levels. Nationality, too, emerged as a helpful predictor, with North American and Western European participants rating stimuli as overall more arousing than Eastern Europeans. Finally, higher scores on the PANAS negative scale translated into stimuli being rated as more arousing, whereas higher scores on the TAS-20 Difficulty Identifying Feeling scale, associated with lower levels of Arousal. The AffectButton version of this model included some of the same predictors, although some were absent, and others were new: cluster membership (again, with the mildly negative cluster equivalent to the neutral one, stimulus modality (with films and images, but not sounds, more arousing than words), baseline Valence and Dominance both being negatively related to the perception of Arousal (i.e., the more positive the mood and confident the attitude at the onset of the study, the less arousing stimuli seemed), and

ITC engagement (positively related to the outcome).

When predicting Dominance as measured with the SAM, all clusters differed significantly from the neutral cluster. In addition, higher baseline Arousal was associated with higher Dominance ratings, and words relative to images were perceived as more conducive to high Dominance. When using the AffectButton, cluster membership and baseline Arousal were still useful as predictors, and they were only joined by baseline Dominance, which was positively associated with the outcome.

Because all the film clips were selected for specific use in this research and did not have any available norms, it was important to assess any potential influence from film-specific covariates, i.e., differences in resolution, bitrate, frame rate and the presence of background music. Interestingly, none of these were associated with PAD ratings, neither on the SAM, nor the AffectButton version of the measures.

At the level of words however, one specific covariate - word frequency - was indeed associated with Arousal ratings, when Arousal was measured with the AffectButton only. Because we found insufficient evidence for the convergent validity of the AffectButton previously, we did not take this as a particular sign of concern, since word frequency did not affect any of the PAD dimensions, when measured using the SAM. Nonetheless, in this same model, higher PHQ depression scores were also associated with overall higher Arousal ratings.

Overall, when differences between the various clusters were taken into account in models, the predictors (still) frequently emerging important were usually those related to ITC items, i.e., engagement, and negative physical symptoms, as well as baseline Arousal.

4.4.7 Limitations

There are some limitations of this study, in particular concerning the cultural differences that sometimes emerged as important in our models. Under ideal circumstances, the sample used should be more homogeneous in terms of nationality / cultural background of participants. However, this can prove to be difficult to achieve in student samples from international study centres.

In addition, ratings for the four ITC-SOPI-SF items will have included some degree of error, since they represented holistic ratings for all the five items included in stimulus blocks, and therefore hinge on how similar these items truly appeared to participants. Perhaps in a simpler design, it would have been possible to request ITC-SOPI-SF ratings for individual stimuli.

Also, it would have been interesting to find a weighted distance measure for the film matching process, and use this to give literature-based weights to the PAD dimensions in the matching process. This could, for instance, have led to Valence and Arousal contributing more to the matching process, and Dominance contributing less - relative to the current solution where all three dimensions are considered equally.

While four types of stimuli were assessed - words, sounds, images and films - only two were measured in terms of perceived presence and engagement - images and films. This is because engagement items often fail to be phrased in general enough terms, to still be applicable to a variety of modalities/media (Lessiter et al., 2001 discuss this issue when proposing the ITC-SOPI items).

Finally, in our models we did not assess any interactions between variables, or polynomial trends (e.g., whether Valence might be non-linearly related to engagement: being higher at very low or very high Valence, but lower for medium Valence), in part because these were not a-priori formulated as key areas of interest, and even for exploratory purposes, would have burdened models and the entire analysis to an excessive degree - particularly in terms of model convergence.

Chapter 5

Study 3B: Introducing non-immersive virtual environments as affective stimuli

5.1 Introduction

THIS study was carried out in order to introduce and explore virtual environments (VEs) as an additional emotion elicitation method, and assess their potential to induce emotional states relative to other stimulus types, e.g., words, sounds, images, and film clips. In order to achieve this, we selected and tested 75 VEs drawn from Second Life, and contrasted them with a set of 25 film clips which we reused from a previous study (see the first assessment of these film clips in Chapter 4).

5.1.1 Aims

As this study constitutes an extension of our previous line of research, we began by verifying whether the current Study 3B sample is similar to that used in Study 3A on a variety of covariates (with Study 3A having been discussed in Chapter 4, and hereafter referred to as 'the previous study'). If the two samples are similar on these covariate measures, then comparing stimulus ratings between the two different samples would be considered acceptable.

Secondly, because some film stimuli from the previous study were reused in the current work, we will use this opportunity to assess how consistently the same films were rated by two different samples. This would constitute valuable evidence that these stimuli present good reliability properties.

In terms of virtual environments (VEs), we tested 75 Second Life (SL) VEs, and at the end of the data collection period, we assessed how/if they had suffered any dramatic changes over time, in terms of object characteristics and placement, general landscape etc. If they had, then those VEs were excluded from further analyses. Of the VEs that remained, we aimed to match 25 of them to the film clips tested here. A related, secondary aim was to match 25 of these VEs to an equal number of films, words, sounds and images in terms of their PAD scores. This was done in order to achieve a final set of stimuli, which would include five modalities: words, sounds, images, films and virtual environments (VEs) - all present in equal proportions, and matched in terms of Valence, Arousal and Dominance ratings.

We further inspected the clustering structure of VEs, independently of / compared to films, in order to investigate the stability / generalisability of the previous clustering solution: i.e., if four clusters emerged for the VEs, as they did for the other modalities previously assessed in this thesis (i.e., words, sounds, images and films). A second reason for clustering VEs and films alike was to verify the extent to which they sampled the PAD 3D space (i.e., whether this was biased to over-represent specific areas etc).

Finally, we also investigated whether, following the VE filtering process (i.e., the transition from 75 to 25 VEs) and the insertion of PAD ratings as covariates, any differences still remained between films and VEs in terms of presence / engagement and physical symptoms. This was done with the expectation that VEs would score higher on both these measures, compared to films.

5.2 Method

5.2.1 Participants

A sample of $N = 60$ volunteering participants was recruited for this study. The participants were remunerated for their time commitment according to the legal minimum wage at the time of their participation. We used the same measures in this study as those previously described in Study 3A (see Section 4.2.3) - and the sample is described according to these measures in Table 5.1. We have also compared this sample to the Study

3A sample in Section 5.3.1, in terms of age, gender, nationality and other variables.

Table 5.1: Sample description in terms of age, media usage, the PANAS schedule, PHQ, TAS-20, and baseline Valence, Arousal and Dominance. PAD baseline values are broken down by session.

No	Measure	Mean	Trim	Median	SD	Min	Max	Range	Skew	Kurt	SE
1	Age	23.62	22.60	23.00	5.58	18.00	46.00	28.00	1.85	3.73	0.72
2	CompComf	3.60	3.75	4.00	0.72	1.00	4.00	3.00	-1.97	3.74	0.09
3	CompFreq	3.97	4.00	4.00	0.18	3.00	4.00	1.00	-5.07	24.11	0.02
4	FilmFreq	2.60	2.58	3.00	0.76	1.00	4.00	3.00	-0.11	-0.41	0.10
5	GameFreq	1.62	1.58	2.00	1.08	0.00	4.00	4.00	0.31	-0.57	0.14
6	MMORPGFreq	0.43	0.35	0.00	0.59	0.00	2.00	2.00	0.97	-0.11	0.08
7	PhotoFreq	2.68	2.73	3.00	0.97	1.00	4.00	3.00	-0.24	-0.95	0.12
8	SLFreq	0.08	0.00	0.00	0.38	0.00	2.00	2.00	4.45	18.81	0.05
9	VWFreq	0.65	0.46	0.00	0.92	0.00	4.00	4.00	1.89	4.02	0.12
10	PANAS_Neg	6.28	6.12	6.00	2.91	1.00	13.00	12.00	0.44	-0.35	0.38
11	PANAS_Pos	13.68	13.77	14.00	2.36	7.00	18.00	11.00	-0.45	0.10	0.30
12	PHQ_total[a]	5.15	4.65	4.00	3.77	0.00	16.00	16.00	1.14	0.88	0.49
13	TAS20_DIF	7.58	6.96	6.00	6.21	0.00	22.00	22.00	0.74	-0.43	0.80
14	TAS20_DDF	7.57	7.27	7.00	3.06	3.00	17.00	14.00	0.87	0.09	0.39
15	Base V: session 1	5.27	5.35	5.00	1.34	0.00	8.00	8.00	-0.92	2.40	0.17
15	Base V: session 2	5.24	5.22	5.00	0.99	2.00	7.00	5.00	-0.37	0.42	0.13
17	Base A: session 1	3.64	3.61	4.00	1.70	0.00	8.00	8.00	0.14	-0.35	0.22
18	Base A: session 2	3.15	3.24	4.00	1.49	0.00	6.00	6.00	-0.50	-0.62	0.19
19	Base D: session 1	4.46	4.41	4.00	1.45	0.00	8.00	8.00	0.02	0.31	0.19
20	Base D: session 2	4.56	4.49	4.00	1.38	0.00	8.00	8.00	0.23	1.78	0.18

[a] It is worth noting that one participant in the current study presented a PHQ-8 depression score of 20, when an 'acceptable' score is considered to be below 10. This deviation was considered large enough for the participant to be excluded from the sample, before proceeding to further analyses.

Note. **Base A** = Baseline Arousal; **Base D** = Baseline Dominance; **Base V** = Baseline Valence; **CompComf** = level of comfort when using a computer; **CompFreq** = frequency of using a computer; **FilmFreq** = frequency of watching films; **GameFreq** = frequency of gaming; **MMORPGFreq** = Massively Multiplayer Online Role-Playing Games gaming frequency; **PANAS_Neg** = PANAS Negative Affect scale; **PANAS_Pos** = PANAS Positive Affect scale; **PhotoFreq** = frequency of taking photos; **PHQ_total** = PHQ-8 total score; **SLFreq** = frequency of using Second Life; **TAS20_DIF** = TAS-20 Difficulty Describing Feelings subscale; **TAS20_DIF** = TAS-20 Difficulty Identifying Feeling subscale; **VWFreq** = frequency of using virtual worlds.

5.2.2 Materials & stimuli

In this study, we used 25 film clips (selected according to the data and method previously described in Section 4.3.2) and 75 VEs collected from Second Life. All VEs have been listed in Appendix E.1. The VEs chosen from Second Life represent a variety of content, e.g., carnivals and fun fairs, gory crime scenes in mansions, dark caves, hospitals and rainy alleys, a cathedral, a palace, beaches etc.

In choosing these materials, we aimed to represent similar content to the other stimuli used throughout this thesis (images, words, sounds and films), in order to make the comparison between modalities more meaningful. Wherever this was not possible, we instead included a VE which would likely mimic similar *emotional* content to the other modalities, despite the *semantic* content being different. For instance, because we could not find a VE to include a youth showing disgust at finding chewing gum firmly stuck to his shoe (which is present in one of our YouTube film stimuli), we instead found a VE including a very dirty back yard filled with piles of litter and scrap material.

5.2.3 Apparatus & instruments

In this study, we displayed VEs using a program named FireStorm (`v.i686_4.7.2.47295`), an open-source and multi-platform Second Life viewer. In order to make this program more suitable for use in this study, we implemented several changes to the software's `settings.xml` file, as shown in Table 5.2 below, for reproducibility:

Table 5.2: FireStorm settings modified within its `settings.xml` file. The 0 and 1 values translate to TRUE/FALSE booleans. Whenever present in pairs, these refer to the current state of the variable, and whether or not this should persist.

No	FireStorm variable affected:	Values modified to:
1	`SLURLTeleportDirectly`	1 & 1
2	`ShowGroupNoticesTopRight`	0 & 0
3	`IgnoreAllNotifications`	1, 1, 0 & 0
4	`NotificationNonFriendIMOptions`	'noaction'
5	`NotificationObjectIMOptions`	'noaction'
6	`FSSendTypingState`	0 & 0
7	`RenderTrackerBeacon`	0 & 0
8	`FSDisableBeaconAfterTeleport`	1 & 1

This was installed on a dual-boot (Windows 8.1, 64-bit & Ubuntu 14.04) Alienware X51 R2 machine with the specifications listed in Table 5.3 (p. 232). Details on the 'immersive', wide-screen monitor used are also included in the same table, which was compiled using the Speccy program (`v1.31.732`)[1].

In addition, because we had no control over the online worlds in Second Life (and they could suffer unexpected alterations along the duration of the study), the solution for overcoming this obstacle was to instead create inventories for all the items in these worlds, multiple times across the data collection. This was done so that, even if changes were to occur in the virtual stimuli, these could be recorded and later used as covariates. For every week that the data collection was in place, the environments were scanned

[1] https://www.piriform.com/speccy

to capture any changes that might have occurred. This was done twice every week, on Tuesday and Friday evenings (for consistency). Overall, over the course of this study, the selected worlds were scanned 25 times. It was not possible to scan worlds simultaneously while the participants were exploring them, due to the fact that this would require that the researcher's own avatar would be present alongside the participant's, and could be distracting or tedious. However, for the scans to generally be possible, a customised version of the FireStorm Second Life viewer was created, courtesy of Dr. Donal Stewart.

In this custom version of the viewer, the menu options for seeing this inventory while logged into any given Second Life world are found, similarly to the official FireStorm release, under the: `'World' menu` → `'Area search'` → `'Filter' tab`. Given the options for filtering the inventory, we allowed only items of the types: `Physical`[2], `Attachment`[3], and `Temporary`[4], but not `Child Prims`[5], as it was considered sufficient to include just `Parent Prims` in the search. `Neighbouring Regions` were also excluded from the search.

Afterwards, clicking the '`Apply`' button automatically switches to the '`List`' tab. This includes all the items based on the selected search / filtering criteria, and described according to the dimensions selected in the '`Options`' tab. All these dialog boxes are included in sequence, in Figure 5.1 (p. 233), for clarity.

[2] i.e., anything affected by physics, e.g., gravity.

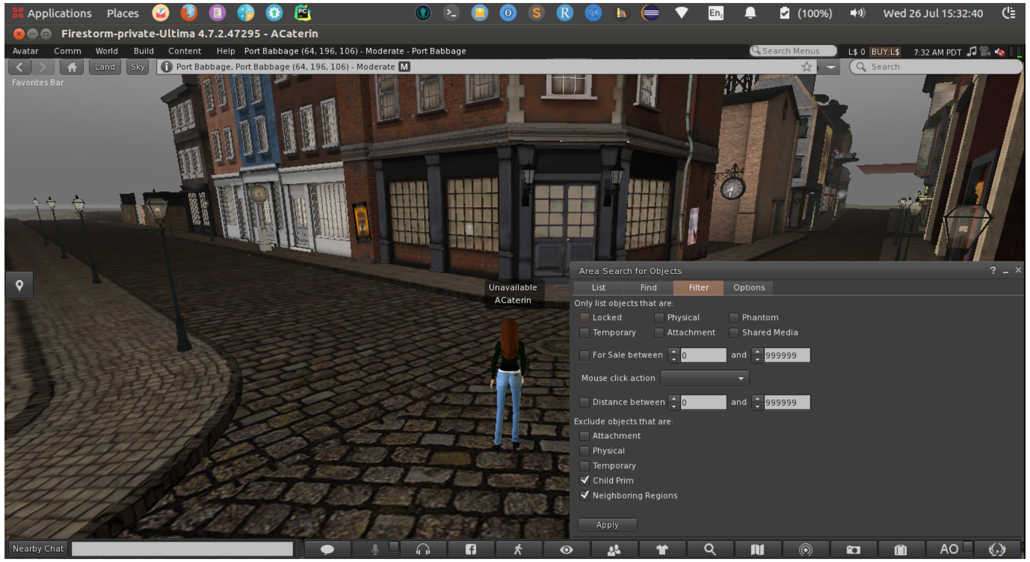
[3] e.g., avatar accessories and clothes.

[4] i.e., objects only allowed to exist for up to approximately one minute in a world, after which they disappear, at which point they may or may not be re-created (according to sources on: `https://community.secondlife.com/forums/topic/83948-temporary-prims-versus-rezzables/`. An example would be any bullets fired and hitting a target, afterwards self-deleting.

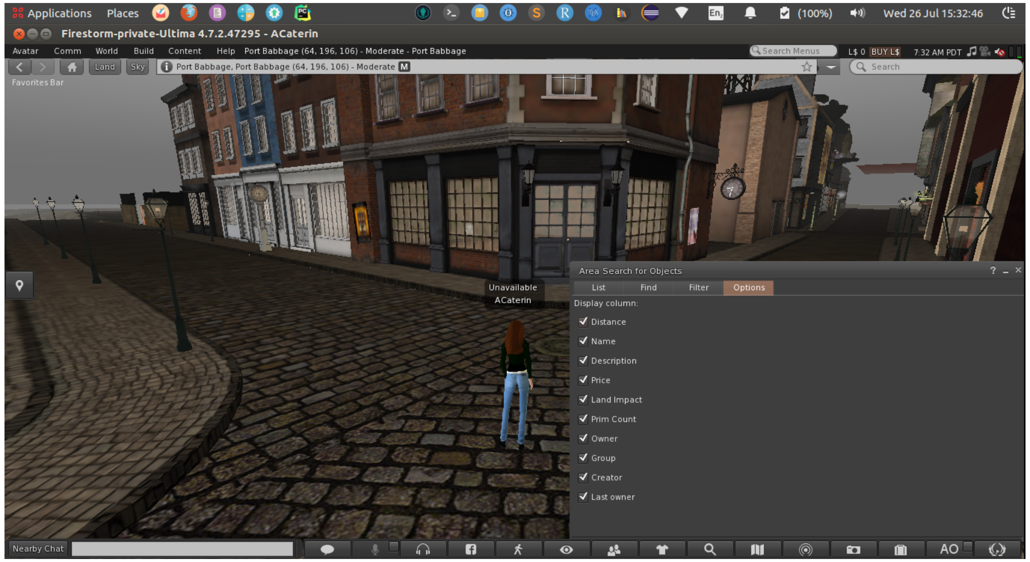
[5] i.e., objects which are components of a larger object, such as the driver's wheel in a car - where the car would be the parent prim.

Table 5.3: Alienware machine and wide-screen monitor used for the VR experiments.

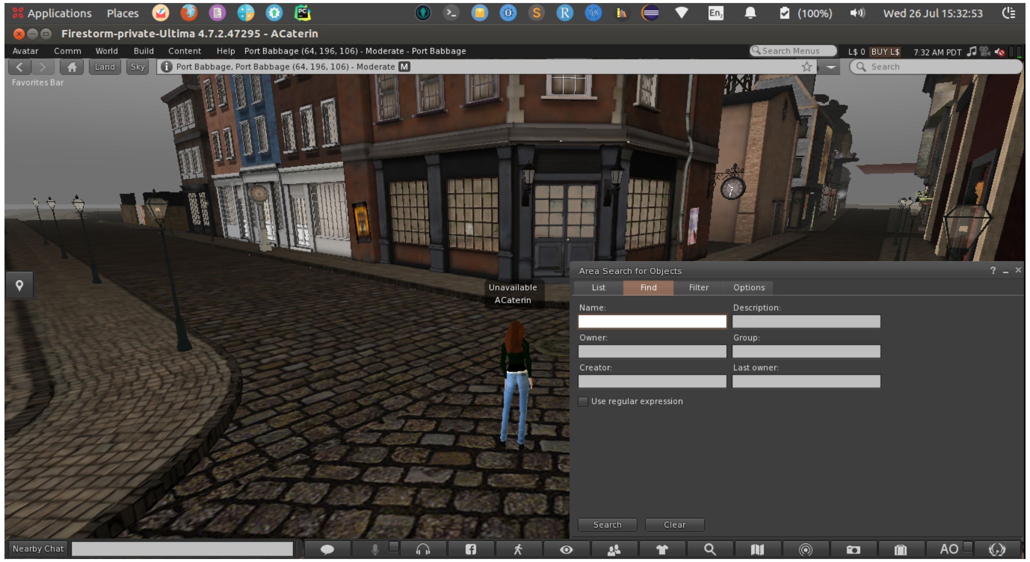
Category	**Subcategory (if applicable)**	**Specifications**
Processor		Intel(R) Xeon(R) processor E3-1200 v3/4th Gen Core processor PCI Express x16 Controller - 0C01
		Intel Core i7-4790 CPU @ 3.60GHz
	Cores	4
	Threads	8
	Microarchitecture	Haswell
	Package	Socket 1150 LGA
Motherboard		Alienware 0PGRP5 (SOCKET 0)
RAM		16.0GB Dual-Channel DDR3 @ 799MHz (11-11-11-28)
	Type	DDR3
	Size	16384 MB
	Channels	Dual
	DRAM Frequency	799.9 MHz
	Physical Memory	16 GB (total, with 13 GB available)
Graphics		Intel HD Graphics 4600 (Dell)
		2047MB NVIDIA GeForce GTX 960
	ForceWare version	368.81 (Drivers)
Monitor		Dell U2410
	Resolution	1920 x 1200 pixels
	Frequency	59 Hz



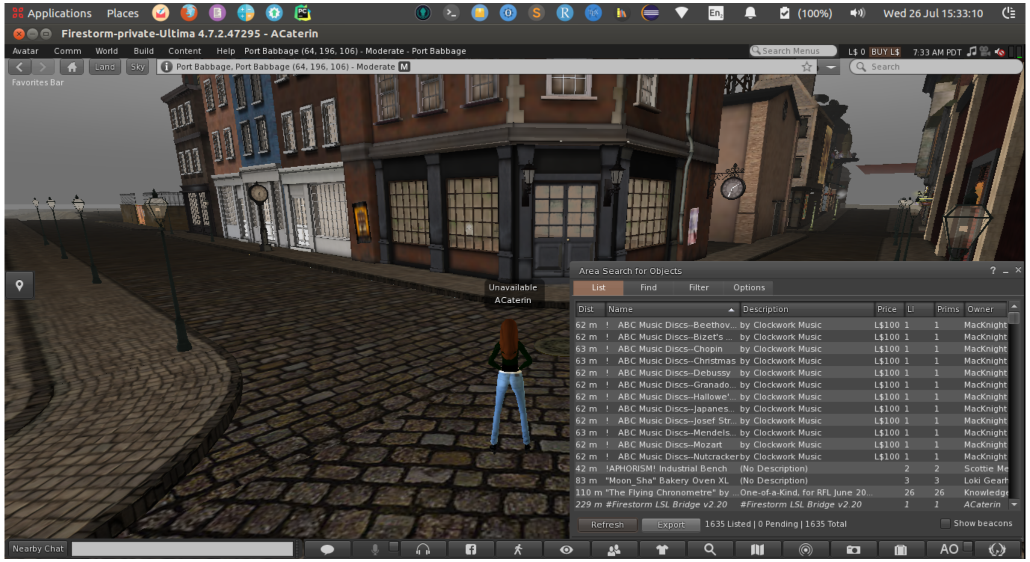
(a) Area scan dialogue: Filter tab



(b) Area scan dialogue: Options tab



(c) Area scan dialogue: Find tab



(d) Area scan dialogue: List tab

Figure 5.1: Required settings to reproduce the area scans executed bi-weekly for this study.

For ease of use, these bi-weekly scans were largely automated using SikuliX (version 1.1.0)[6] - a very powerful piece of software which uses image recognition to identify and control graphical user interface (GUI) components. SikuliX can also control the system mouse and keyboard, in order to interact with the GUI. In other words, taking snapshots of buttons and menus and presenting them to SikuliX in the sequence in which they needed to be activated, allowed this software to emulate user behaviour, and trigger repetitive actions on its own, i.e., clicking on specific menus and typing specific input into text fields etc. In the case of Second Life, the SikuliX workflow is shown in Figures 5.2 and 5.3 and emulates the menu navigation described in the previous paragraphs (i.e., `'World' menu` → `'Area search'` → `'Filter' tab`), followed by a loop through all the VEs used - each of which would be typed by SikuliX into FireStorm's navigation bar. After waiting 100s for the object inventory to become populated, SikuliX would refresh the list, and then export it as a `.csv` file before moving on to the next VE.

[6] Downloadable from: http://sikulix.com/

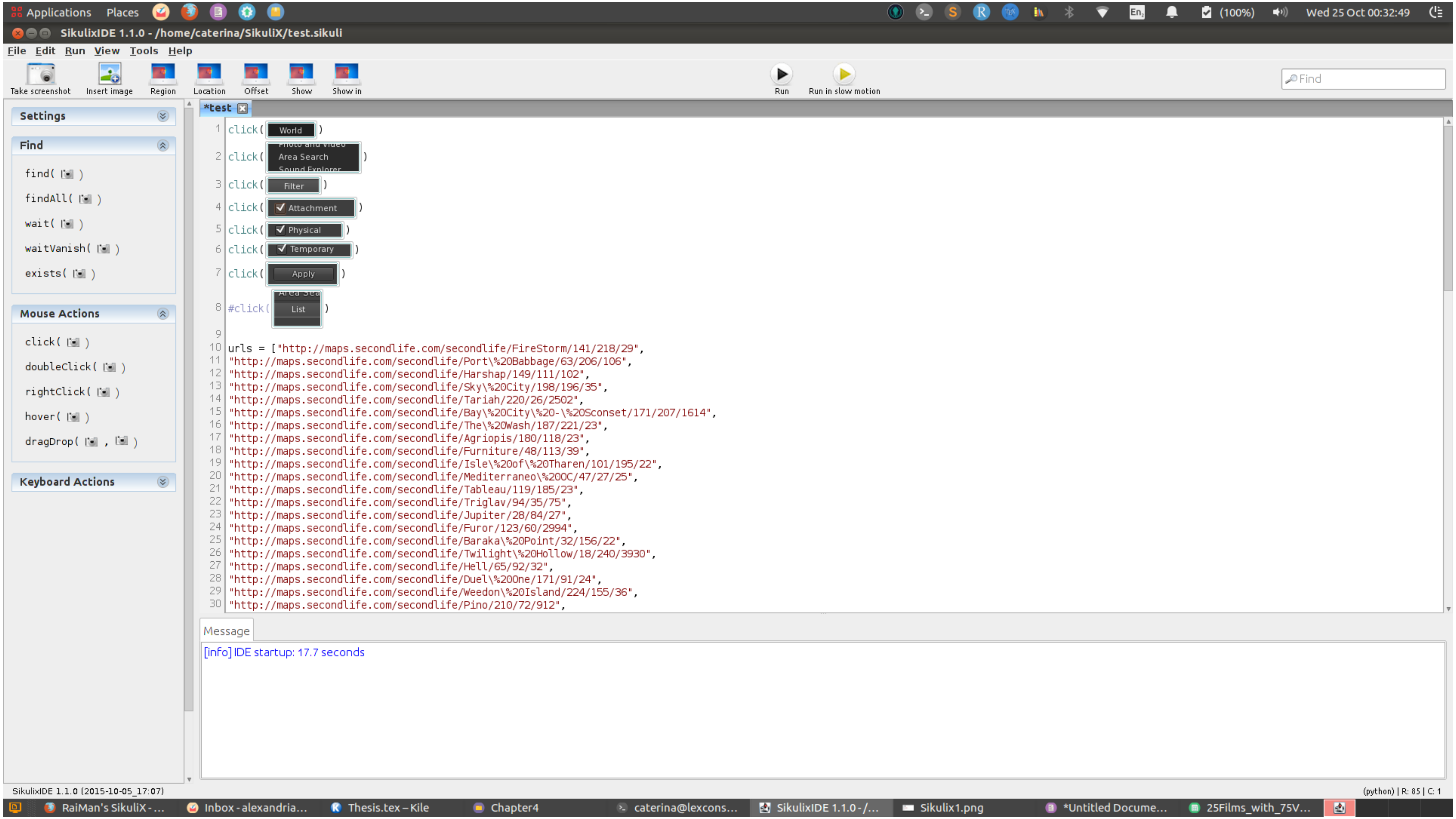


Figure 5.2: (1) SikuliX workflow for scanning VEs twice weekly.

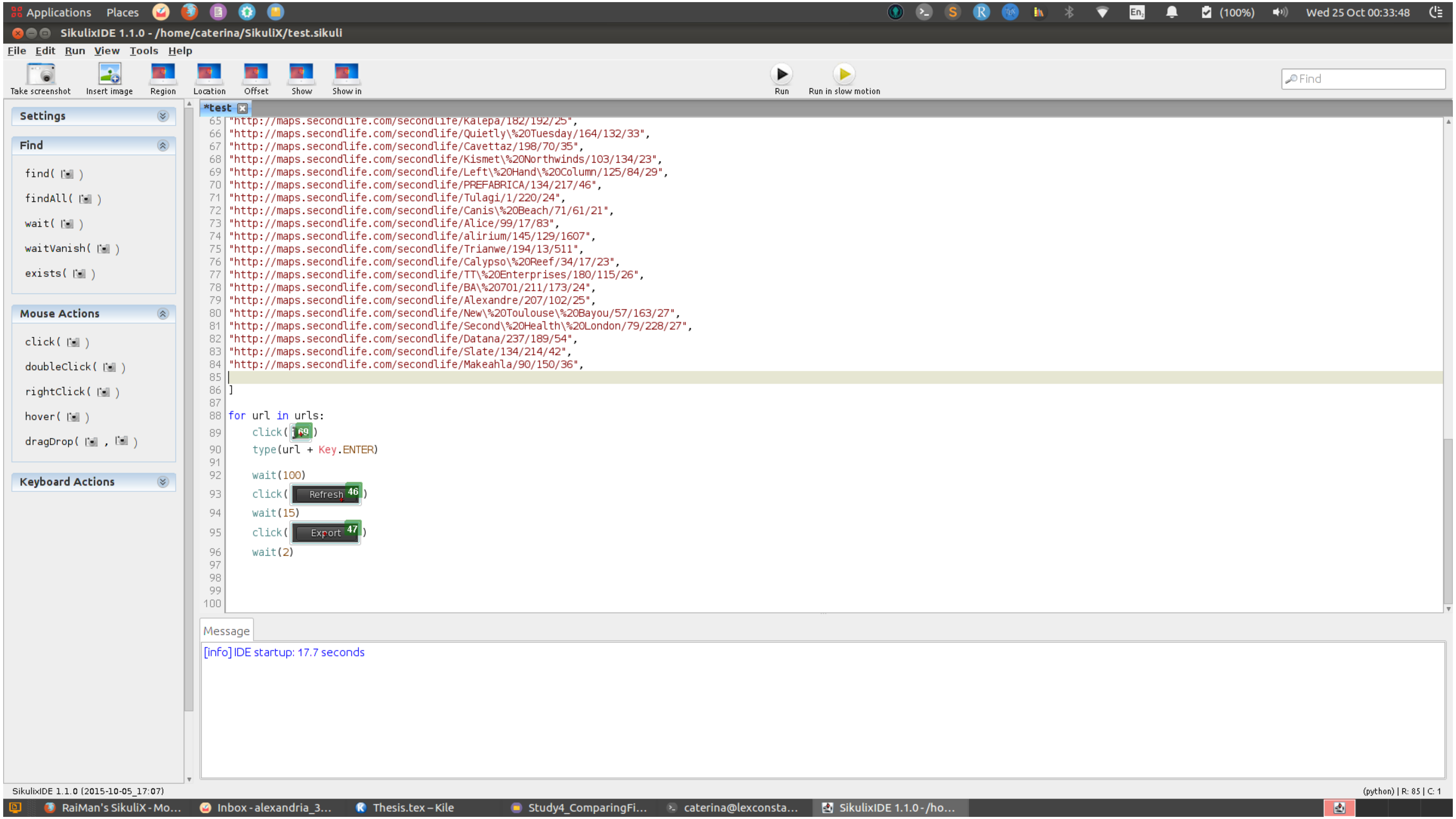


Figure 5.3: (2) SikuliX workflow for scanning VEs twice weekly.

Finally, it was also necessary to fix the time of day in these environments, for experimental control purposes - since this normally fluctuates based on any settings built in by the owner of the world. FireStorm allows users to modify this via the: `‘World’ menu` → `‘Environment editor’` → `‘Environment Settings’ dialog`. There, users can set the same time of day for all environments, as shown in Figure 5.4 (p. 237).

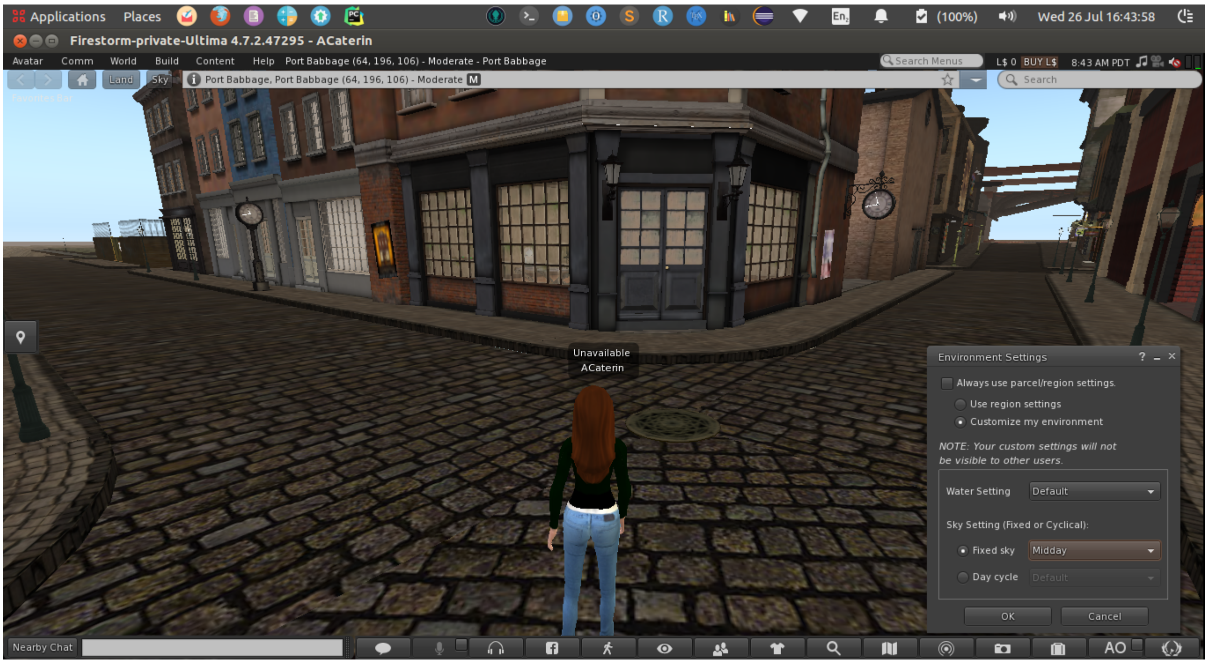


Figure 5.4: Required settings to fix the time of day in Second Life.

5.2.4 Design

This study was constructed using a within-subjects, repeated-measures design, where each participant was able to view and rate both the film clips and the VEs in random order, and across two study sessions (given the large number of stimuli to view and rate). In line with the previous experiments in this thesis, the stimuli for each modality were organised into clusters, where each cluster was represented by multiple, smaller stimulus ‘blocks’[7]. Each such block contained 5 stimuli of the same modality (i.e., either films or VEs) and drawn from the same cluster. As such, instead of either randomising all the stimuli individually (regardless of their modality), or randomising the stimuli at the level of entire clusters, we randomised the *stimulus blocks* in a way that is consistent with the nested structure of stimuli (i.e., stimuli within blocks, within clusters, which in

[7] Similarly to films, VE stimuli were sampled for inclusion in this study on the basis of intuitively matching pre-existing sounds, images and words (SIW). Selecting films and VEs was done after the SIW stimuli had already been clustered by themselves, and hence, the search for which film or VE stimulus to include in this study was done separately for each SIW cluster, and while bearing in mind the general meaning of that cluster

turn are nested within modalities - see Appendix E.1 for details).

The advantage of this randomisation scheme was that residual amounts of affect carrying over to the next stimulus and affecting ratings are less likely, e.g. a pleasant image followed by a negative film might reduce the negativity rating for the latter. In contrast, seeing 5 negative films in a row followed by 5 pleasant images in a row would attenuate this effect. Despite the possibility of habituation effects occurring instead (e.g., with ratings to 5 negative stimuli regressing towards the mean with prolonged exposure), this should be mitigated by the fact that, within each block, the stimuli are presented in random order. As such, at the level of the entire sample, bias produced by habituation should not affect any particular stimuli to any overwhelming degree. A second benefit for this randomisation scheme was the possibility of measuring presence / interactivity variables at the end of each block, rather than after each individual stimulus, which would have been too time-consuming.

5.2.5 Procedure

Upon presenting themselves to the research lab, participants were invited to read an instruction sheet generally describing the experimental procedure, and explaining how to complete the rating tasks[8] and navigate through the VEs (e.g., being informed that clicking on doors in Second Life will open them). Afterwards, they were able to put this information to use within a practise phase of the experiment in OpenSesame, which included the exploration and rating of 3 example VEs. Finally, once this was complete, the experimenter left the testing room and allowed participants to carry on with the study by themselves.

In the main experiment, participants began their first study session by providing ratings for their baseline emotional state (using three 9-point SAM scales, as well as the AffectButton), and answering the same questionnaire measures as shown in Chapter 4. They subsequently repeated the PAD rating procedure in relation to each stimulus they were presented with: either film clips to view, or VEs to explore in Second Life.

VEs were available for exploration for 1.5 minutes, and film clips were approximately 40s long. This discrepancy was set in place because the amount of information conveyed by watching a very coherent video excerpt in 40s would likely exceed the amount of information gathered by navigating a less systematic virtual expanse. Because typically the online VEs required a certain amount of time to load, participants had to wait for 10s in front of a black OpenSesame screen before having the VE revealed to them and

[8] The SAM scales were explained to participants using the same terminology employed by Bradley & Lang, 1994 when labelling scale ends, and the images used to display the SAM within Opensesame were taken from PXLab (Irtel, 2008; Suk, 2006; Suk & Irtel, 2010) - who in turn inspired their computerised version from the original work by Bradley and Lang (1994); Lang (1980).

ready for exploration. Once the 10s had elapsed, a Python script within OpenSesame would call on `wmctrl` (a window manager control program for Linux) to shift window focus from OpenSesame to Second Life. This revealed the randomly selected VE for that given trial, which would simultaneously set off the timer for the 1.5 minutes afforded for VE exploration. The end of this period was marked by an alarm clock sound, which played for 2s and faded out. Immediately afterwards, `wmctrl` would be called again to shift window focus back onto OpenSesame, which requested ratings from the participant.

In the case of film clips, these were immediately ready to view, and varied somewhat in terms of: whether or not they included any music (which is also true of VEs), their video bitrate, their resolution, precise duration and the number of frames per second. As such, all these variables were logged for later use as covariates, in case they were shown to affect participant ratings.

As discussed and implemented previously, this design saw the stimuli organised into blocks of five, all nested within a given cluster. Because the ratio of films to VEs was `1:3`, this lead to one block of *films* and three of *VEs* being included into every cluster. In addition, the overall number of 25 VEs assigned to each cluster were randomly divided across these smaller blocks of five. Once all stimuli were divvied up between blocks, the blocks themselves were randomly assigned to one of two experimental sessions lasting roughly one hour each - with session order being counter-balanced between participants.

Several settings were set in place to ensure the VE exploration experience was as smooth as possible for participants. For instance, they were logged in Second Life as either male or female avatars, to match their real genders. The online status for participants was set to 'Unavailable' to limit interactions with other individuals / avatars, and the movement speed through the VEs was set to 'Running' (as 'Walking' speed realistically could not provide an adequate coverage of the virtual worlds, given that the presentation time for the VEs was 1.5 minutes). In addition, the camera view was set to first person, in order to mimic the viewing conditions for, and facilitate comparisons with, the upcoming head-mounted display (HMD) study, where by definition a first-person viewing experience would be implemented. Finally, the microphone was muted, so that no other Second Life users could hear any sounds from the researcher or the participant.

Due to the fact that prolonged use of an HMD (i.e., the Oculus Rift DK1) can lead to nausea, and also considering that it can be both time consuming and tiring to switch frequently between an HMD and a computer screen, for this study we displayed the 75 virtual Second Life stimuli on a large computer screen only.

Finally, during the whole procedure, participants' desktop activity was filmed (only the screen itself, not the individual participants or the surroundings), firstly to make sure they were following the study indications correctly, and as a means for explaining any unusual values in the collected data.

5.3 Results

All results were computed in R, with package information presented in Appendix Section E.4.

5.3.1 Sample comparison between Study 3A and Study 3B

The previous study included 25 words, 25 sounds, 25 images and 75 film clips, whereas the current study includes a sub-selection of these film clips, as well as one new modality: virtual environments. Because the intention was to compare all these modalities against each other, the data from both studies will be pooled together for this to be possible. However, before this can be achieved, we first verified whether the participant samples were similar enough on a set of covariates, to allow for comparison - see Tables 5.4 and 5.5.

Table 5.4: Describing the Study 3A and Study 3B samples in terms of nationality.

Source	Study1	Study2
Nationality: Eastern Europe	16	9
Nationality: Far East	13	11
Nationality: Middle East	1	3
Nationality: North America	5	7
Nationality: South America	1	1
Nationality: Western Europe	24	25
Nationality: Africa	0	1
Nationality; Australia	0	2

Table 5.5: Describing the Study 3A and Study 3B samples in terms of gender.

Source	Study1	Study2
Gender: Female	30	30
Gender: Male	30	29

Other comparisons between samples, based on continuous measures are presented below, in Tables 5.6 and 5.7 (p. 241 - 242).

Table 5.6: Comparing continuous, single-measure, covariates in Study 3A vs Study 3B. The table represents t-test results when determining whether there were differences between the two samples. No significant differences emerged.

No	Measure	Δ	μ_1	μ_2	t	p	df	CI lower	CI upper
1	Age	1.04	23.62	22.58	1.15	0.25	110.16	-0.76	2.84
2	CompComf	-0.11	3.60	3.71	-1.02	0.31	100.30	-0.33	0.11
3	CompFreq	-0.03	3.97	4.00	-1.43	0.16	59.00	-0.08	0.01
4	FilmFreq	0.06	2.60	2.54	0.41	0.68	116.90	-0.22	0.34
5	GameFreq	0.06	1.62	1.56	0.27	0.78	114.98	-0.36	0.47
6	MMORPGFreq	-0.21	0.43	0.64	-1.40	0.16	94.21	-0.51	0.09
7	PhotoFreq	-0.06	2.68	2.75	-0.33	0.74	114.76	-0.44	0.31
8	SLFreq	0.02	0.08	0.07	0.24	0.81	113.54	-0.11	0.14
9	VWFreq	0.11	0.65	0.54	0.68	0.50	115.86	-0.21	0.42
10	PANAS_Neg	0.25	6.28	6.03	0.46	0.65	116.61	-0.83	1.33
11	PANAS_Pos	0.92	13.68	12.76	1.55	0.13	94.72	-0.26	2.10
12	PHQ_total	-0.48	5.15	5.63	-0.72	0.47	116.50	-1.79	0.84
13	TAS20_DIF	0.46	7.58	7.12	0.43	0.67	115.80	-1.67	2.60
14	TAS20_DDF	-0.03	7.57	7.59	-0.05	0.96	115.62	-1.07	1.02

Note. **Base A** = Baseline Arousal; **Base D** = Baseline Dominance; **Base V** = Baseline Valence; **CompComf** = level of comfort when using a computer; **CompFreq** = frequency of using a computer; **FilmFreq** = frequency of watching films; **GameFreq** = frequency of gaming; **MMORPGFreq** = Massively Multiplayer Online Role-Playing Games gaming frequency; **PANAS_Neg** = PANAS Negative Affect scale; **PANAS_Pos** = PANAS Positive Affect scale; **PhotoFreq** = frequency of taking photos; **PHQ_total** = PHQ-8 total score; **SLFreq** = frequency of using Second Life; **TAS20_DIF** = TAS-20 Difficulty Describing Feelings subscale; **TAS20_DIF** = TAS-20 Difficulty Identifying Feeling subscale; **VWFreq** = frequency of using virtual worlds.

Table 5.7: Comparing continuous, repeated-measure, covariates in Study 3A vs Study 3B: PAD baselines. Comparisons were carried out by session, e.g., for the first row in this table, the comparison refers to session 1 baseline measure of Valence, in Study 3A vs. Study 3B.

Comparison	Session	Measure	Δ between studies	μ_1	μ_2	t	p	df	CI lower	CI upper
1	1	Base V	0.42	5.63	5.22	1.81	0.07	112.83	-0.04	0.87
2	1	Base A	0.05	3.63	3.58	0.16	0.88	117.99	-0.58	0.68
3	1	Base D	-0.08	4.32	4.40	-0.33	0.74	113.32	-0.58	0.41
4	2	Base V	-0.05	5.18	5.23	-0.26	0.80	114.91	-0.44	0.34
5	2	Base A	0.05	3.20	3.15	0.17	0.86	115.57	-0.53	0.63
6	2	Base D	-0.58	3.97	4.55	-2.39	**0.02**	117.68	-1.07	-0.10

Overall, these results confirm the very high degree of similarity between independent samples, with the single exception of Dominance baselines recorded during the second session in Study 3A vs. Study 3B, which did differ significantly between samples. This is also underlined, in graphical format, in Figure 5.5 (p. 242).

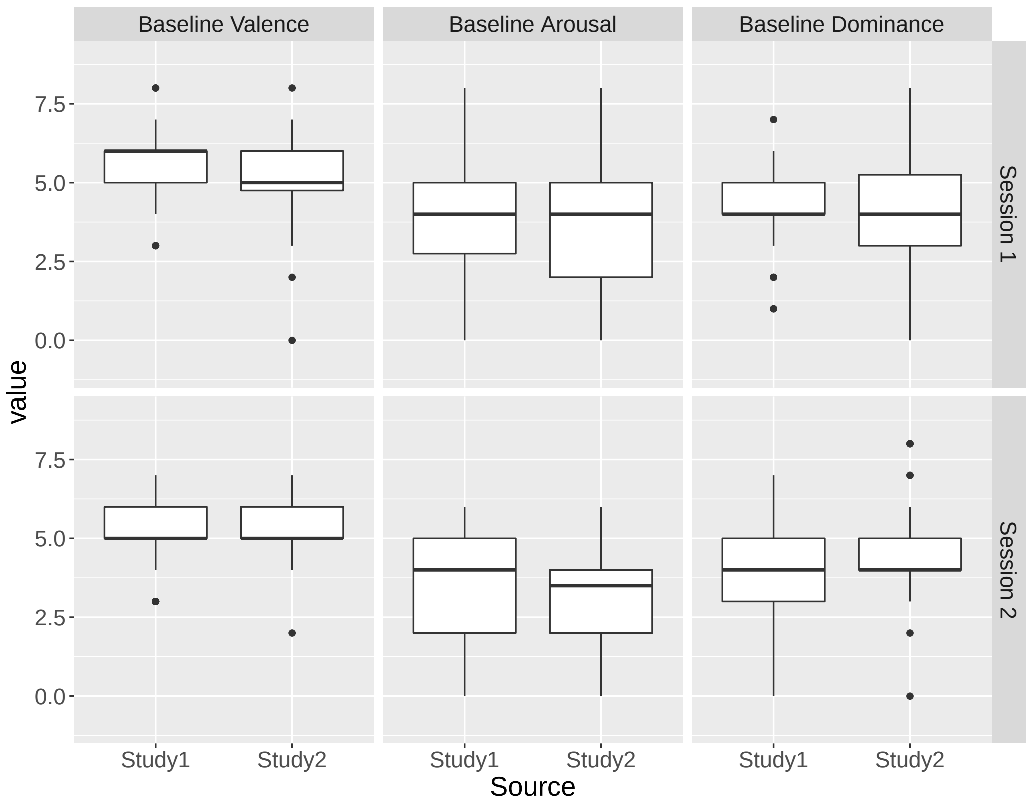


Figure 5.5: Comparing PAD baselines between Studies 3A and 3B.

5.3.2 Comparing film ratings between Study 3A and Study 3B

Interestingly, not only are the covariates very similar overall between studies, but the 25 films used in both Study 3A and Study 3B were rated almost identically by the different participants, as can be seen from Figure 5.6 (p. 244). In fact, correlating the averaged PAD ratings[9] per film between the two studies leads to extremely high Pearson r values: $r = 0.987$ for Valence, $r = 0.975$ for Arousal and $r = 0.950$ for Dominance.

5.3.3 Assessing VE changes over time and selecting consistent VEs

In order to compare VEs to films, it was first necessary to exclude any VEs which changed too extensively over the course of the study. To achieve this, a method had to be devised to quantify the amount of VE change over time, based on repeated environment scans / item inventories, which could be collected twice a week (see Section 5.2.3, p. 230).

We met various obstacles in trying to achieve this. For instance, the Linden Lab programming language, which underlies SecondLife, does not allow scanning of a radius larger than 96m around the point of observation[10], in a virtual world. This meant that either very large areas could not be assessed for changes (any objects over 96m further away from the point of teleportation / entry to a given world, in any direction, would be ignored), or that whole worlds would have to be split into sub-areas smaller than the distance limitation, which could then be scanned sequentially. The latter option was attempted with Full-SIM Object Scanner Pro (Boxed)[11] - a proprietary (although low-cost) Second Life world scanner. A known bug occurred while using this software, and in the absence of any support from the authors, it was abandoned in favour of the customised FireStorm viewer previously described in Section 5.2.3.

Using the custom FireStorm Area Search to assess changes within each VE over time, it was first necessary to select one scan date as a 'baseline', or 'reference point', and to use this as a comparison term for every other scan date (i.e., test dates). While this reference point could naturally have been the start date of the study, we in fact cycled through all the 25 scan dates available, and used each in turn as a reference point, to see if any/which may optimise the level of similarity across inventory scans. This strategy was implemented in order to preserve as many VEs as possible, and observe how VEs changed, not just with reference to the start of the study, but also along its entire duration.

After selecting a given date/scan as a reference or baseline, our algorithm then cycled through each VE, and compared its recorded inventory on the reference date, to its

[9] As collected with the SAM scales, due to the relatively poor performance of the AffectButton.

[10] Limitation inherent in using the Linden Labs `llSensor` function for Second Life.

[11] Full-SIM Object Scanner Pro (Boxed) can be downloaded from https://marketplace.secondlife.com/p/Full-SIM-Object-Scanner-Pro-BOXED/2428734.

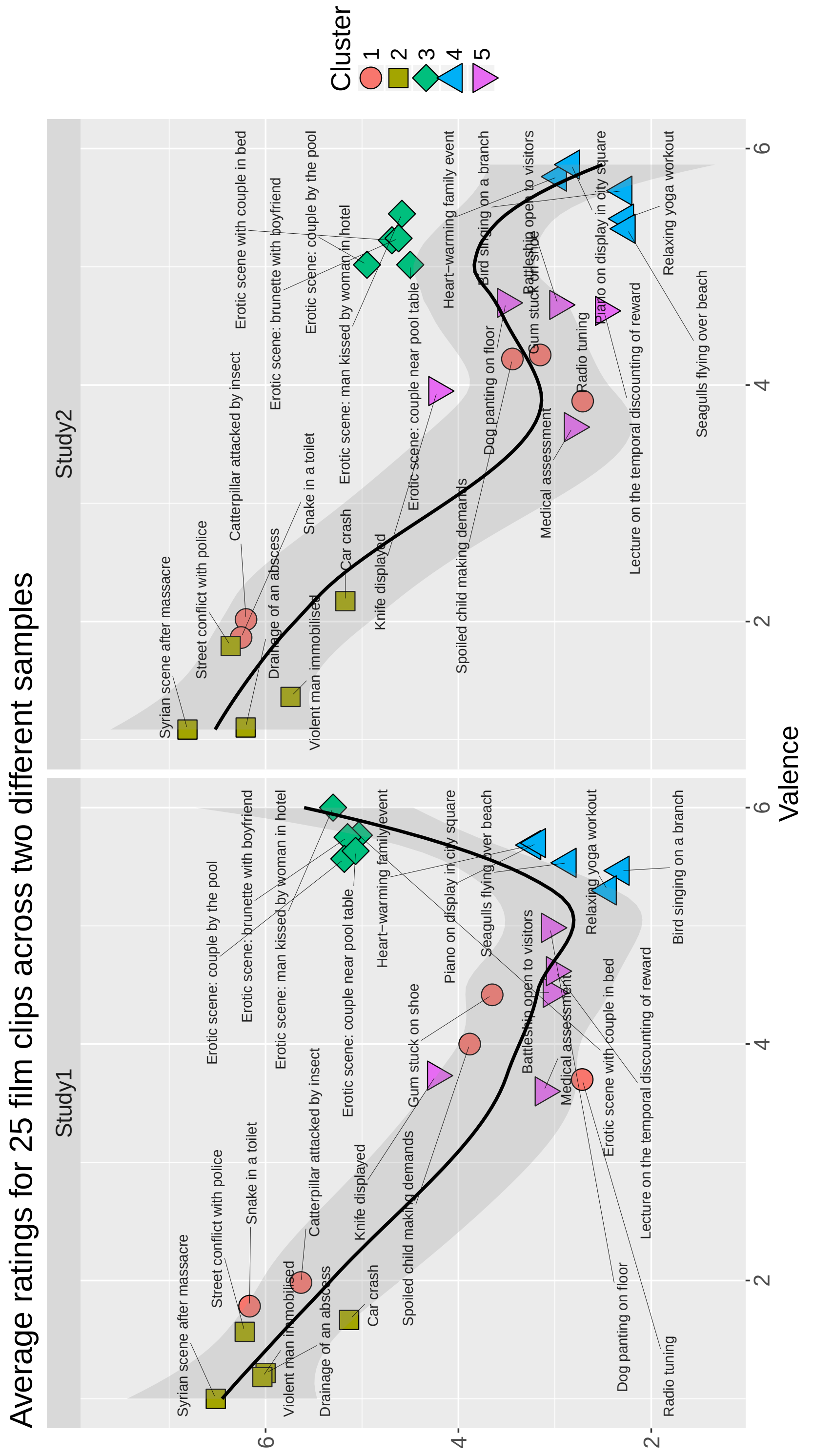


Figure 5.6: Comparing average PAD ratings for 25 films clips used in both Study 3A and 3B.

inventory on each of the remaining test dates. The degree of similarity between the two was computed by first measuring how many items were identical across both the reference date and the test date[12], and how many were unique (i.e., the number of reference items that had vanished, and/or new items that had appeared in the test date). Once these two values were derived, the level of similarity was calculated simply as:

$$S_{VE} = \left(\frac{R}{T}\right) \cdot 2, \; where: \tag{5.1}$$

S_{VE} = the level of similarity between a given VE's inventory on a reference vs. a test date; R = the number of recurring/identical elements across the two scan dates; T = the total number of inventory items found across the two dates, whether identical or unique.

Above, the ratio is multiplied by 2 in case the inventory stayed identical between the reference and the test date. In such cases, were they to exist, the number of total items (T) would simply be the double of the items found in common between scans (R), and therefore, the ratio would only equal 0.5. For ease of interpretation, we preferred a result where 1 corresponds to complete similarity, hence the multiplication by 2. Results of these calculations were stored as a two-tiered list in `R`: we nested all the reference vs. test scan comparisons within a specific VE, and all VEs were, in turn, nested under a particular reference-point.

Based on this information (i.e., by-VE, and by-reference-date comparisons between inventory lists), we relied on a simple global measure for selecting one 'optimal' reference date[13]: grand means. Firstly, for every reference date and VE combination, we computed the average level of similarity across all *test-dates*. For instance, for the VE region named 'Agriopis', and with a selected reference date of 2015-11-03 (i.e., the starting date of the study), similarity had dropped to 0.86[14] by the time the next scan occurred (i.e., 2015-11-06), and then dropped again (dramatically) to 0.28 by the following test-date after that (i.e., 2015-11-11). By the last test-date (i.e., 2016-02-05), the average level of similarity across all test-dates for 'Agriopis' was 0.77, with an SD = 0.15. These calculations were performed separately for all reference dates and all VEs, and then used to create grand means for every reference date. It is this information that is presented below in Figure 5.7 (p. 246), where the location of points represents the average similarity recorded across all VEs for a given reference date, and the size of the points represents the SD across the within-VE similarities (i.e., the SD for values that contributed to the by-reference-point grand means).

The 12^{th} scan (2015-12-08) proved to be the optimal choice for a reference scan. When using it as such, the similarities derived for all the test dates are shown using a

[12] Inventory items were considered to re-occur between the reference and the test scan, if they presented

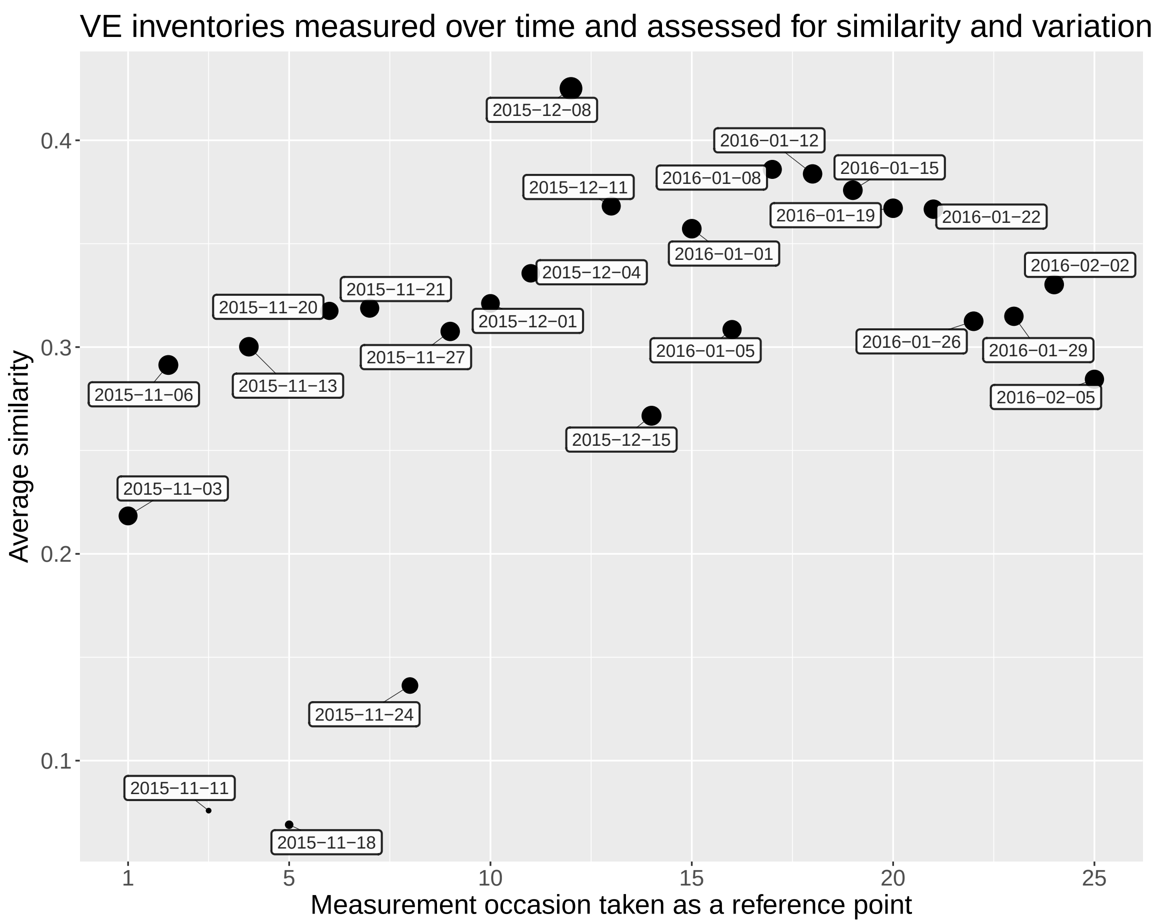


Figure 5.7: Grand means for VE similarity, for each reference date: similarities were first averaged by VE, within a reference date. Afterwards, these means (i.e., single values for each VE) were averaged into a global value per reference date. The highest average similarity is thus achieved when the 12^{th} scan (2015-12-08) is used as a reference point. While the 5^{th} measurement / scan (2015-11-18) presented the lowest amount of variability in similarity scores, it was also simultaneously the worst performer in terms of average similarity.

heatmap, in Figure 5.8 (p. 247). Here it is already fairly clear which VEs were consistent in terms of their inventories, and which, less so.

an identical `Name`, and `X, Y, Z` coordinates. See Figure 5.1 (c), p. 233, for all other variables recorded.

[13] i.e., which maximised the average within-VE similarity levels (across test-dates).

[14] As calculated with the short formula from 5.1.

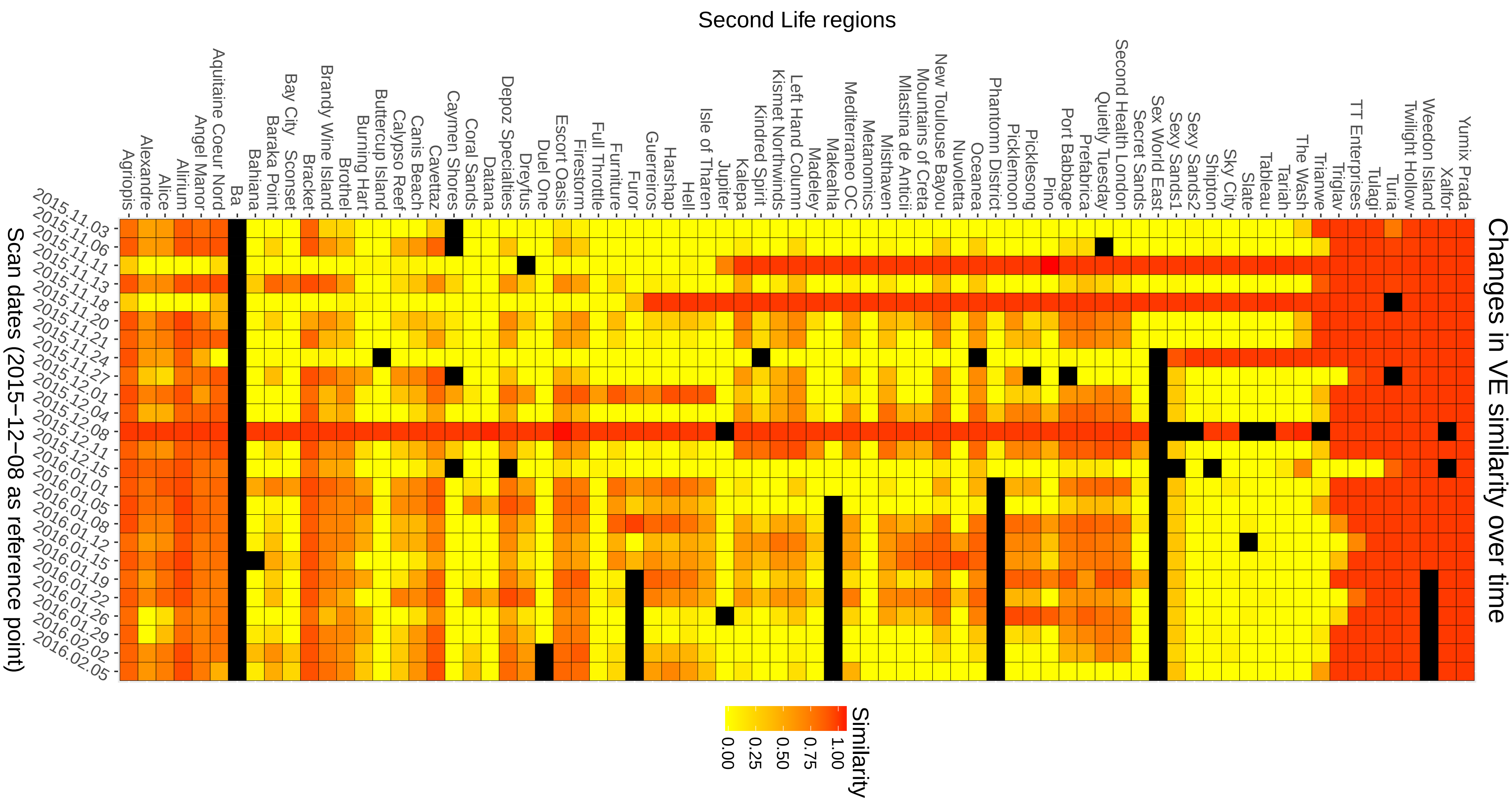


Figure 5.8: Heatmap of VE similarity tracked over time. Because scans on 2015-12-08 were taken as a reference point against which to compare all test scans, the associated column represents the overlap between the reference point and itself. Hence, this column presents maximum similarity and should be ignored. Black cells represent missing values (i.e., a scan could not be completed on that date because the VE was inaccessible in Second Life).

With the 12^{th} scan date used as a reference, the optimised similarities which we generated for all VEs were used in a model-based cluster analysis, in order to exclude the VEs which changed too dramatically over the course of the study. More specifically, for each VE, we computed the average similarity achieved during the study, as well as the variability (SD) in these measures of similarity, over the 25 scans. These two dimensions were used in a cluster analysis in order to find a group of highly inconsistent VEs (i.e., with low average similarity, and high variability). Based on these data, five clusters were proposed, of which only one could qualify as highly inconsistent. The parameters / centroids for each cluster are presented in Table 5.8 (alongside the associated Figure 5.9, p. 249), with the third cluster appearing as the best candidate for exclusion.

Table 5.8: Parameters / centroids, for each 'consistency' cluster.

Dimension	Cluster 1 (N = 9)	Cluster 2 (N = 14)	Cluster 3 (N = 23)	Cluster 4 (N = 24)	Cluster 5 (N = 4)
Means	0.84	0.58	0.15	0.35	1.00
SDs	0.23	0.31	0.32	0.35	0.00

Cluster 3 was indeed excluded, leading to a number of 52 remaining VEs for the rest of this analysis. In addition, one further VE ('BA') disappeared immediately from the onset of the study (see earlier heatmap), hence this VE was also excluded, leaving a total of 51 VEs. For all these remaining VEs, we noted their similarities for each scanning date, and merged them with participant data, based on which participant session was closest to which VE scan date[15]. This was done so that the similarities for the remaining 51 VEs could still function as covariates later in this analysis. Results of this nearest-date matching process are illustrated in Figure 5.10 (p. 250).

5.3.4 Feasibility of matching VEs to other stimulus modalities

Before embarking on another matching procedure using the `R` `optmatch` package (see Chapter 3, p. 99, and Chapter 4, p. 149), we investigated the differences and/or similarities between the PAD distributions of VEs and the other modalities used thus far - see Figures 5.11 (p. 251) and 5.12 (p. 252). It emerged that a matching procedure would have poor chances of success, because the PAD ranges for the 51 VEs were severely restricted compared to other modalities, especially images. We were able to investigate this by pooling the data from Study 3A (i.e., the source of word, sound and image ratings), with the data from Study 3B, for film and VE ratings. Films rated in Study 3A were not considered here simply because they were shown to be extremely correlated to how those same film clips were rated in Study 3B, as mentioned above.

[15] using a so-called 'rolling-join' via the `R` `data.table` package.

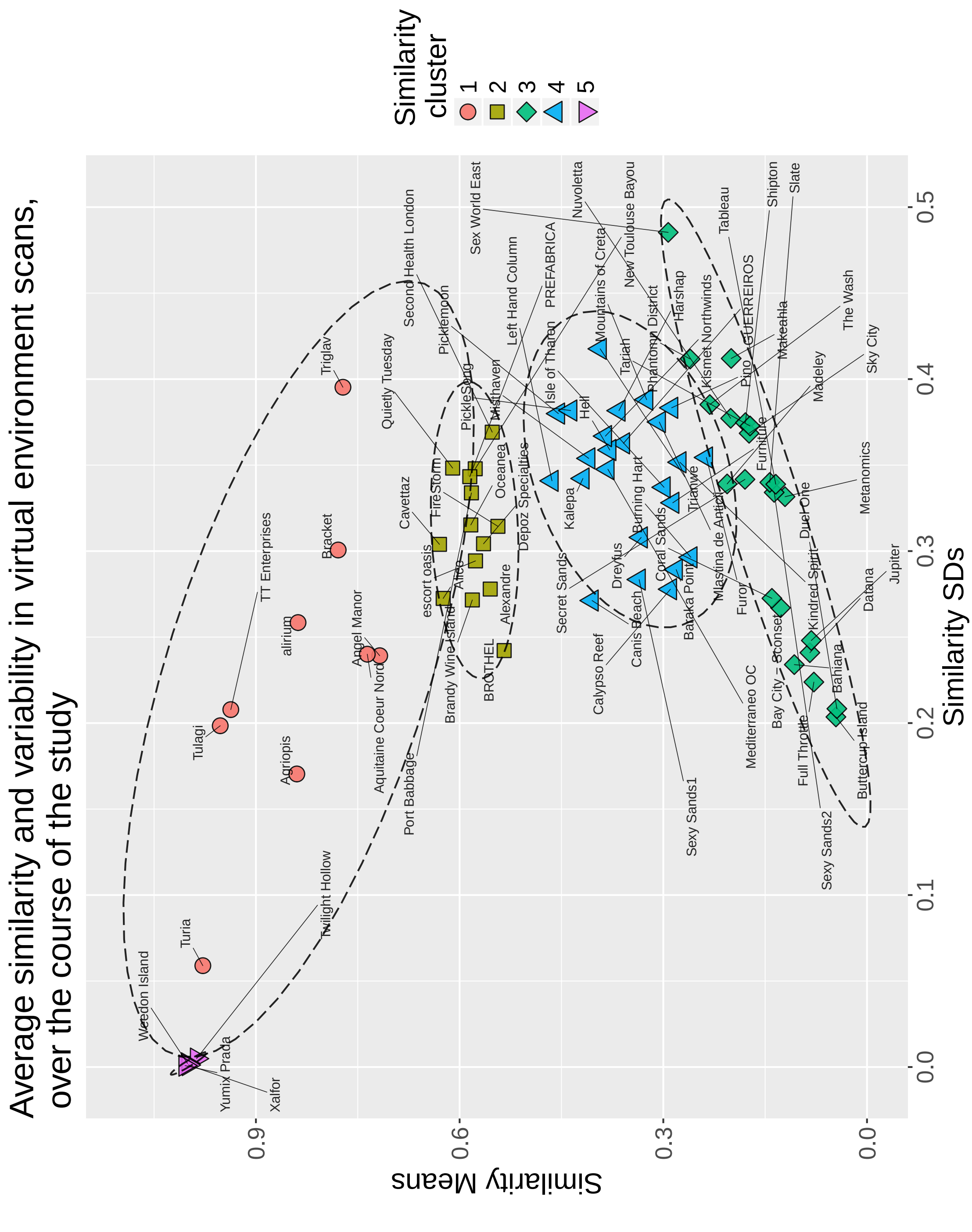


Figure 5.9: Assessing VE similarity over the course of the study, in order to remove the most inconsistent group of VEs.

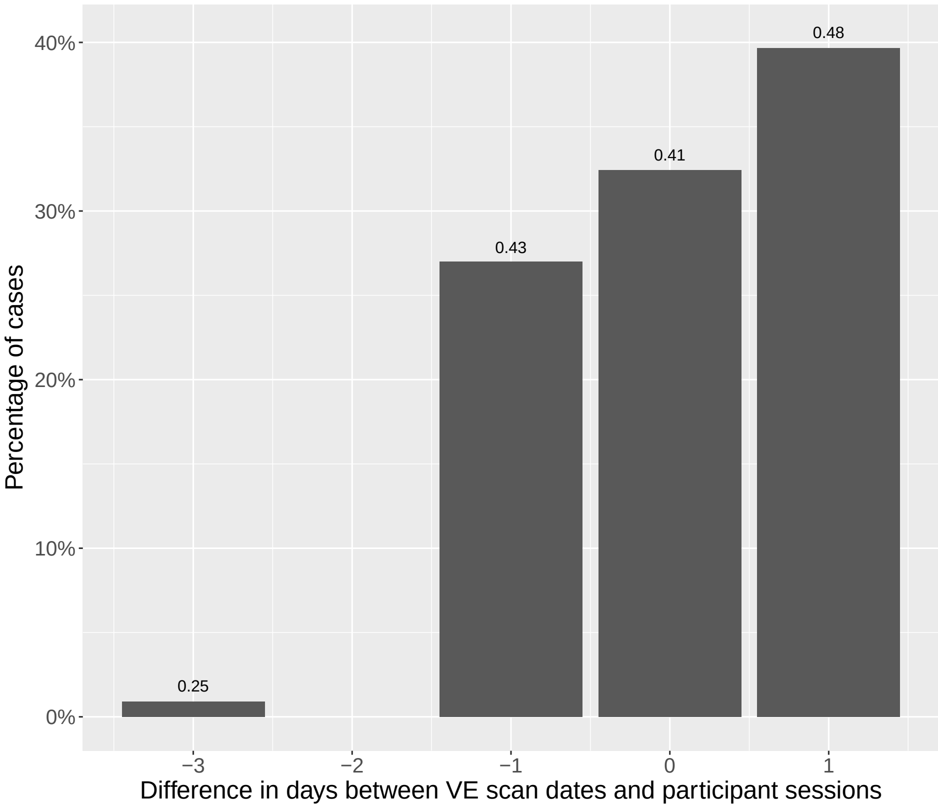


Figure 5.10: Nearest-date matching of VE inventory scans to participant sessions. Usually, the VE scan dates happened to occur within a day of participant sessions (before or after), or actually on the same day (although scans always occurred later in the evening). Regardless, this tended to not affect the average similarity recorded for participants. Rarely, and at the very beginning of the study, VE scans pre-dated participant sessions by 3 days, when average similarity was highest.

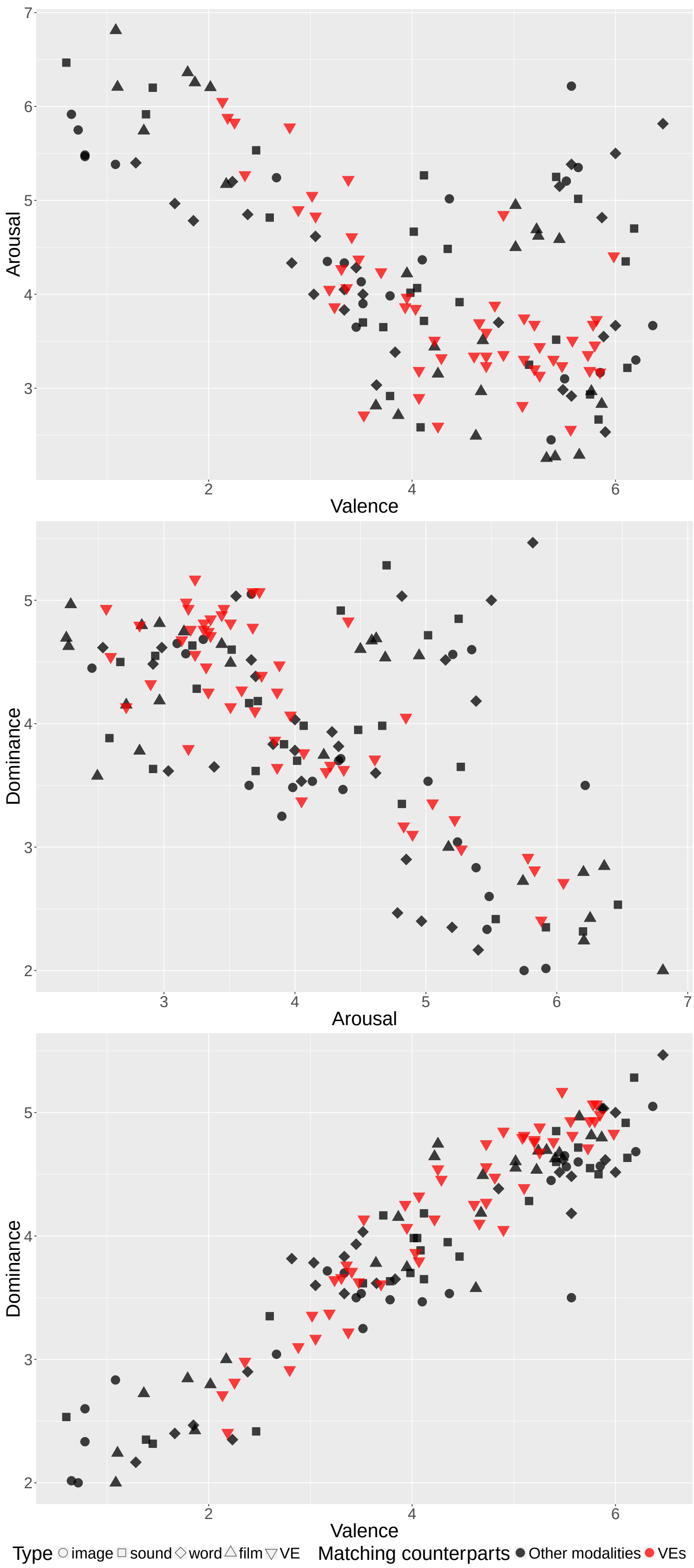


Figure 5.11: Multiple bivariate scatterplots, precluding a matching process between VEs and other modalities - with VEs marked using downward facing red triangles.

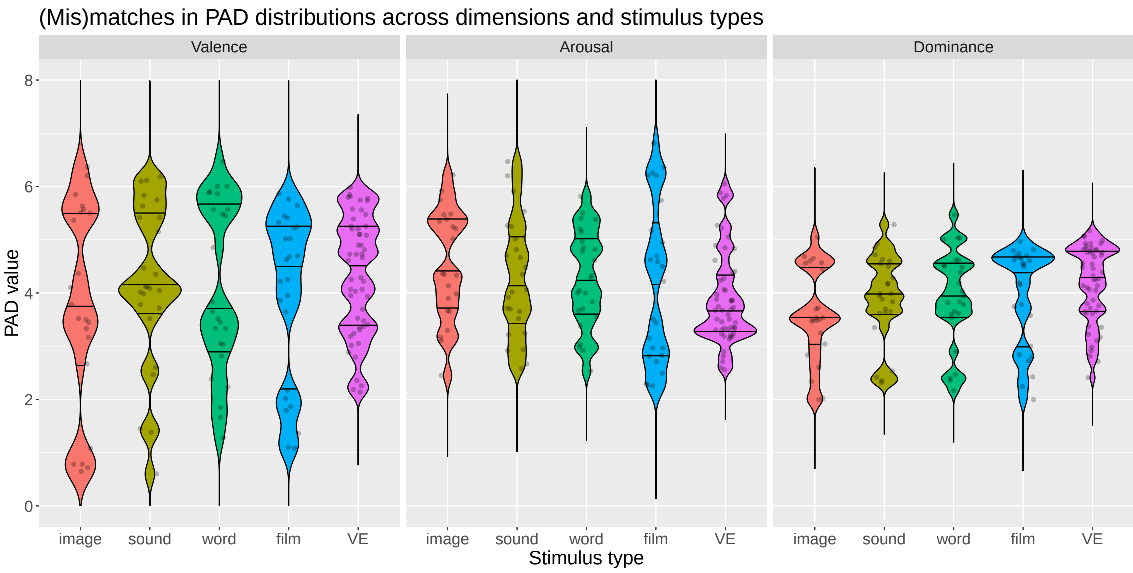


Figure 5.12: Violin plots, with jittered data points printed underneath, also precluding a matching process between VEs and other modalities: Valence range restriction is very obvious for VEs compared with images, but also the other modalities illustrated. Horizontal lines represent quartiles.

5.3.5 Clustering structure for VEs

Because a direct, data-driven method for matching stimuli 1:1 across modalities did not show promise, we opted for a different approach to selecting a subset of 25 of the 75 VEs: running another cluster analysis and extracting stimuli from the clusters based on their uncertainties. This option also presents the added benefit of giving insight into the dimensionality and structure of this VE dataset.

After averaging the raw PAD ratings received by each VE, we fed this aggregated data to a model-based cluster analysis, which identified only 2 components / clusters[16], with ellipsoidal shapes, equal volume and orientation (EVE model), and BIC = -228.85 (see Table 5.9, and Figure 5.13, p. 255 for cluster centroids and sizes, and an illustration of group membership and uncertainties, respectively). The top 3 best-fitting models according to the BIC criterion (all of them occurring for $k = 2$, and models: EVE, VVE and EVV) are signalled **in bold** in Table 5.10 (p. 254), where rows express each value of k between 1 and 9, and columns represent all the possible model types (EII, VII etc.)[17].

Table 5.9: VE-only cluster centroids and sizes, for $k = 2$. These seem to follow a coarser pattern of general Positive Affect (with above-average Valence, relatively low Arousal and close-to-moderate Dominance) and Negative Affect (with lower Valence, higher Arousal than the positive cluster, and lower Dominance than the positive cluster), instead of the finer-grained categories discovered for images, sounds, words and film clips.

Dimension	Cluster 1 ($N = 30$)	Cluster 2 ($N = 21$)
Valence	5.11	3.25
Arousal	3.41	4.56
Dominance	4.66	3.46

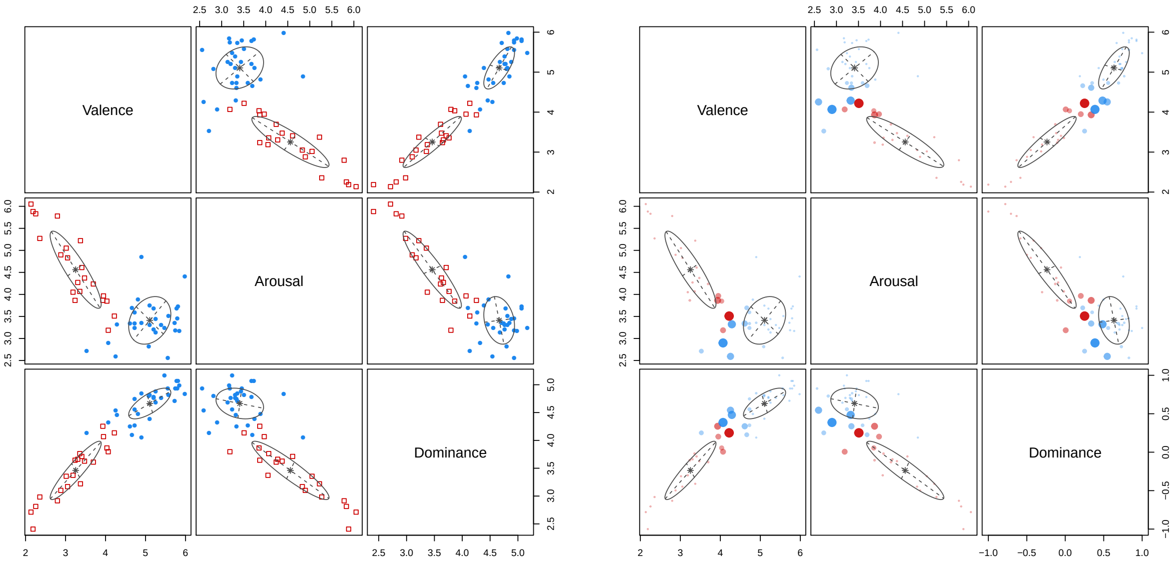
As discussed in previous chapters, MBC analysis also provides cases with a level of uncertainty for having been assigned to a certain cluster. These uncertainties are listed below, where the 51 VEs are also described in terms of the cluster they were assigned to, regardless of (or given) a specific value for uncertainty - see Table 5.11 (p. 256). The top VEs with the lowest uncertainty from each cluster are also marked in bold, given these were sampled for use with a head-mounted display (HMD) in the following study (see Study 3C, Chapter 6, p. 271).

[16] Perhaps unsurprisingly, given the restricted PAD range for this modality.

[17] To foreshadow an upcoming analysis, the average BIC value for the expected solutions showed poor fit here: for the $k = \mathbf{5}$ solutions, the average BIC was -289.275 (SD = 17.280), and for the $k = \mathbf{4}$ solution, the average BIC = -274.310, with SD = 13.913. These values indeed suggest a much poorer fit compared to the $k = 2$ EVE model (where, as a reminder, BIC = -228.85).

Table 5.10: BIC values for MBC models, upon clustering exclusively VEs in this study. The top three best-fitting models are signalled with a bold font.

k	EII	VII	EEI	VEI	EVI	VVI	EEE	EVE	VEE	VVE	EEV	VEV	EVV	VVV
1	-420.855	-420.855	-419.536	-419.536	-419.536	-419.536	-240.062	-240.062	-240.062	-240.062	-240.062	-240.062	-240.062	-240.062
2	-328.067	-331.539	-323.838	-326.6	-325.794	-329.299	-239.674	**-228.85**	-240.981	**-232.733**	-241.144	-243.749	**-236.941**	-240.487
3	-303.223	-313.537	-291.837	-295.788	-305.622	-306.769	-241.599	-243.39	-247.871	-245.85	-253.6	-258.73	-258.748	-262.189
4	-289.028	-291.694	-269.775	-274.773	-287.19	-290.355	-248.935	-264.745	-253.733	-262.919	-265.455	-274.416	-286.076	-281.251
5	-290.58	-299.486	-277.49	-291.66	-302.757	-314.286	-258.782	NA	NA	NA	-279.156	NA	NA	NA
6	-299.635	-311.114	-282.92	-296.984	-316.181	-320.896	-266.157	NA	NA	NA	-304.494	-321.106	NA	NA
7	-293.058	-302.902	-282.62	-289.251	-321.452	-322.483	-278.337	NA	NA	NA	-329.241	-330.234	NA	NA
8	-295.731	-312.706	-287.219	-305.803	-322.692	-341.493	-276.057	NA	NA	NA	-336.406	-354.585	NA	NA
9	-304.812	NA	-308.762	NA	NA	NA	-286.249	NA	NA	NA	-360.791	NA	NA	NA



(a) Stimulus classification into the 2 clusters.

(b) Stimulus membership uncertainty: the larger and darker the points, the higher the uncertainty.

Figure 5.13: The two VE clusters created using model-based clustering.

Table 5.11: Uncertainties for $k = 2$ MBC model, upon clustering exclusively VEs. Within each cluster, the VEs are sorted by uncertainty.

No	Second Life Region	Valence	Arousal	Dominance	Uncertainty	Cluster
1	**Kalepa**	5.983	4.407	4.831	0.000000	1
2	**Yumix Prada**	5.814	3.729	5.068	0.000000	1
3	**Brandy Wine Island**	5.780	3.678	5.068	0.000000	1
4	Angel Manor	5.797	3.458	4.932	0.000000	1
5	Mountains of Creta	5.847	3.169	4.983	0.000000	1
6	Kismet Northwinds	5.576	3.508	4.814	0.000000	1
7	Quietly Tuesday	5.729	3.356	4.712	0.000000	1
8	Oceanea	5.746	3.186	4.932	0.000000	1
9	Pino	4.898	4.847	4.051	0.000000	1
10	Depoz Specialties	5.475	3.237	5.169	0.000001	1
11	New Toulouse Bayou	5.203	3.678	4.780	0.000003	1
12	Canis Beach	5.254	3.441	4.881	0.000009	1
13	Burning Hart	5.390	3.305	4.763	0.000011	1
14	Alexandre	5.102	3.746	4.390	0.000050	1
15	TT Enterprises	5.203	3.203	4.763	0.000231	1
16	Bracket	5.102	3.305	4.814	0.000233	1
17	Aquitaine Coeur Nord	5.254	3.136	4.678	0.000368	1
18	Calypso Reef	4.814	3.881	4.475	0.000379	1
19	Misthaven	4.898	3.356	4.847	0.001174	1
20	alirium	5.559	2.559	4.932	0.001561	1
21	Alice	4.729	3.339	4.746	0.008327	1
22	Cavettaz	5.085	2.814	4.797	0.009178	1
23	Port Babbage	4.729	3.593	4.271	0.017222	1
24	Left Hand Column	4.729	3.237	4.559	0.024156	1
25	Triglav	4.661	3.695	4.102	0.051412	1
26	Trianwe	3.525	2.712	4.136	0.052174	1
27	Mediterraneo OC	4.610	3.339	4.254	0.122179	1
28	Furniture	4.254	2.593	4.542	0.189003	1
29	Dreyfus	4.288	3.322	4.458	0.275215	1
30	Second Health London	4.068	2.898	4.322	0.370516	1
31	**Twilight Hollow**	2.186	5.881	2.407	0.000000	2
32	**Picklemoon**	2.136	6.051	2.712	0.000000	2
33	**Hell**	2.254	5.831	2.814	0.000000	2
34	**PickleSong**	2.797	5.780	2.915	0.000000	2
35	Baraka Point	2.356	5.271	2.983	0.000000	2
36	Harshap	2.881	4.898	3.102	0.000000	2
37	Weedon Island	3.017	5.051	3.356	0.000000	2
38	Furor	3.051	4.831	3.169	0.000000	2
39	Mlastina de Anticii	3.373	5.220	3.220	0.000001	2
40	Escort Oasis	3.407	4.610	3.712	0.000135	2
41	Prefabrica	3.186	4.051	3.373	0.000142	2
42	Brothel	3.305	4.271	3.661	0.000215	2

43	Guerreiros	3.475	4.373	3.627	0.000285	2
44	FireStorm	3.356	4.068	3.763	0.001133	2
45	Tulagi	3.695	4.237	3.610	0.001378	2
46	Isle of Tharen	3.237	3.864	3.644	0.001957	2
47	Xalfor	4.034	3.847	3.864	0.064103	2
48	Sexy Sands1	3.949	3.966	4.068	0.099205	2
49	Turia	4.068	3.186	3.797	0.104171	2
50	Agriopis	3.932	3.864	4.254	0.170348	2
51	Kindred Spirit	4.220	3.508	4.136	0.403547	2

Interestingly, these results deviate considerably from the number of clusters expected originally, $k = \mathbf{5}$ (i.e., 4 data-driven and 1 'artificial' neutral cluster), which emerged for *IAPS images* at the beginning of this work (see Section 2.3.2, p. 75, and Table 2.2, p. 81), and which helped in planning all the subsequent research. Because all additional modalities (i.e., words, sounds, films and VEs) were introduced after IAPS images and with the intention to match them, these added modalities copied the same (intended) cluster structure that had been derived for IAPS images. The mismatch between this intended five-cluster structure (modelled on images), and the two empirical VE clusters, is cross-tabulated in Table 5.12.

Table 5.12: Crosstabulation between intended, 5-cluster classification structure, and the observed 2-cluster structure for VE ratings. The table shows little convergence between the two classifications, with the exception of the fourth *'intended'* cluster (i.e., positive and serene items), which is entirely part of the general positive *observed* cluster.

	Intended cluster				
Observed cluster	1	2	3	4	5
1	4	2	2	13	9
2	5	8	6	0	2

Be that as it may, we also forced a $k = 5$ MBC model in order to see if, based on these empirical data, the classification bore any similarity to our own intended clusters, or still differs widely. Figure 5.14 (p. 258) illustrates this new classification, and Table 5.13 (p. 258) contains the cluster centroids and sizes for this solution.

Unfortunately, VEs largely do not respect the expected clustering structure, as indicated by the heatmap / crosstabs in Figure 5.15 (p. 260) - again with the exception of the fourth positive and serene cluster, which is largely intact and is retrieved by the MBC solution as the second cluster in the data-driven solution. Overlap indices confirm this pattern of very modest convergence, with the Jaccard and Adjusted Rand Indices equalling 0.272 and 0.265, respectively.

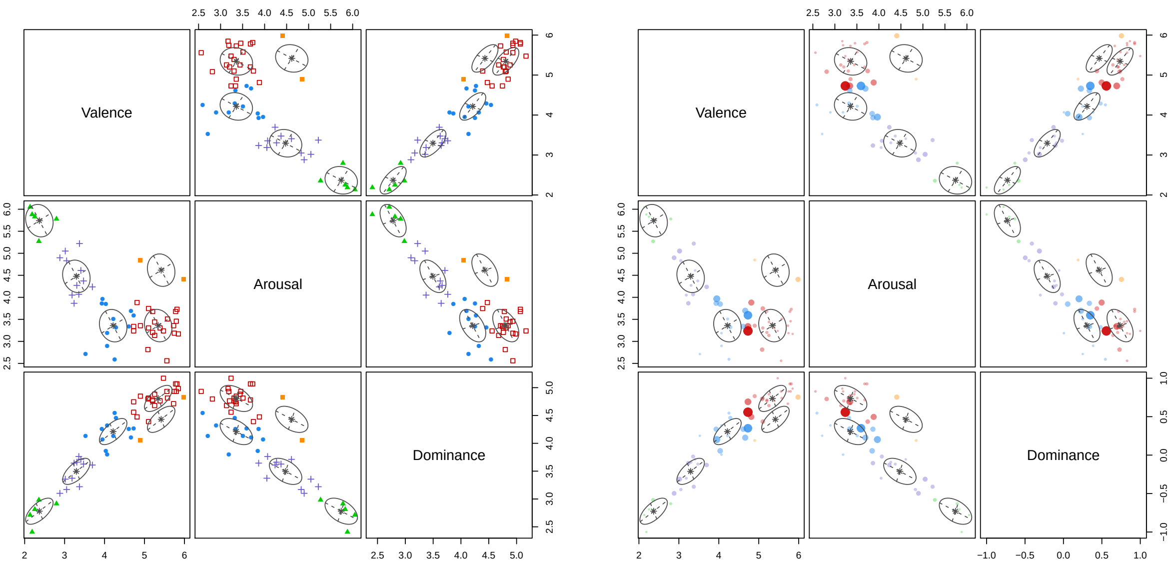
Table 5.13: Centroids resulting from a forced classification of VEs into 5 clusters.

Dimension	Cluster 1 ($N = 12$)	Cluster 2 ($N = 21$)	Cluster 3 ($N = 5$)	Cluster 4 ($N = 11$)	Cluster 5 ($N = 2$)
Valence	4.22	5.34	2.37	3.29	5.42
Arousal	3.36	3.36	5.74	4.48	4.62
Dominance	4.21	4.81	2.78	3.50	4.43

Based on previous results, another possible clustering outcome could have been $k =$ **4**, based on the number of clusters previously discovered empirically for words, sounds and images matched together (see Section 3.3.3.1, p. 115), and for images, words, sounds *and films* (see Subsection 4.3.4.1, p. 176) - however, this too contradicts the coarse, two-cluster solution supported by the VE data.

5.3.6 Quadratic trends across all modalities

Interestingly, despite VEs not following the same data structure as the other modalities, they do show a similar deviation from a linear relationship between Valence and Arousal. This quadratic trend is illustrated in Figure 5.16 (p. 260).



(a) Forced stimulus classification into 5 clusters.

(b) Stimulus membership uncertainty: the larger and darker the points, the higher the uncertainty.

Figure 5.14: Five VE clusters created using model-based clustering, which interestingly, emphasises the previously discussed U-shape in the Valence-Arousal bivariate space.

5.3.7 Clustering films

Given the large deviation of VEs from the expected clustering solution, we verified whether the same is true of films, when clustered alone based on their average PAD ratings, without the influence of any other modalities mixed in. In this case, MBC analysis led to an optimal cluster number of $k = 3$, for a VEE model (ellipsoidal clusters, with equal shape and orientation). This too represents a deviation from the expected results (i.e., $k = 4$), albeit a minor one, because two $k = 4$ models are also included among the top three best-fitting models. As before, these have been indicated in bold within Table 5.14 (p. 261). These three film clusters presented the following properties, as shown in Table 5.15 (p. 261), and Figure 5.17 (p. 262).

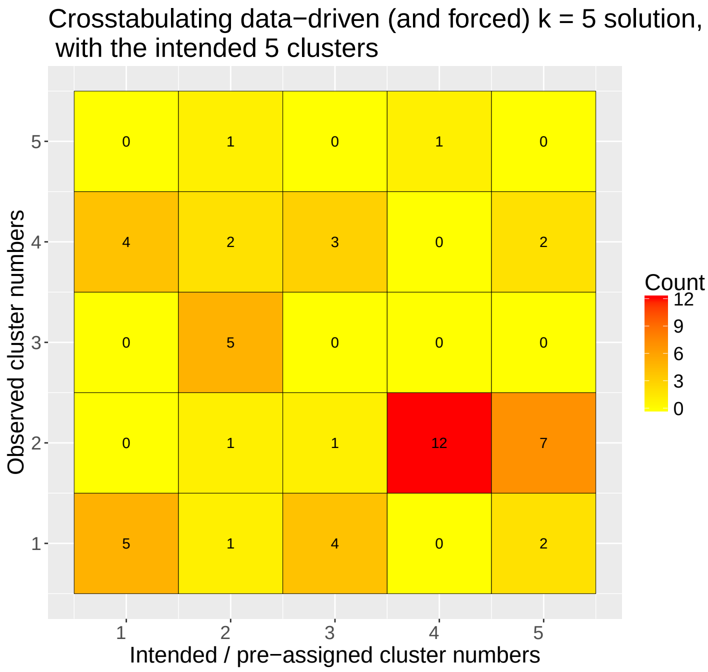


Figure 5.15: Forced $k = 5$ crosstabulation for VEs.

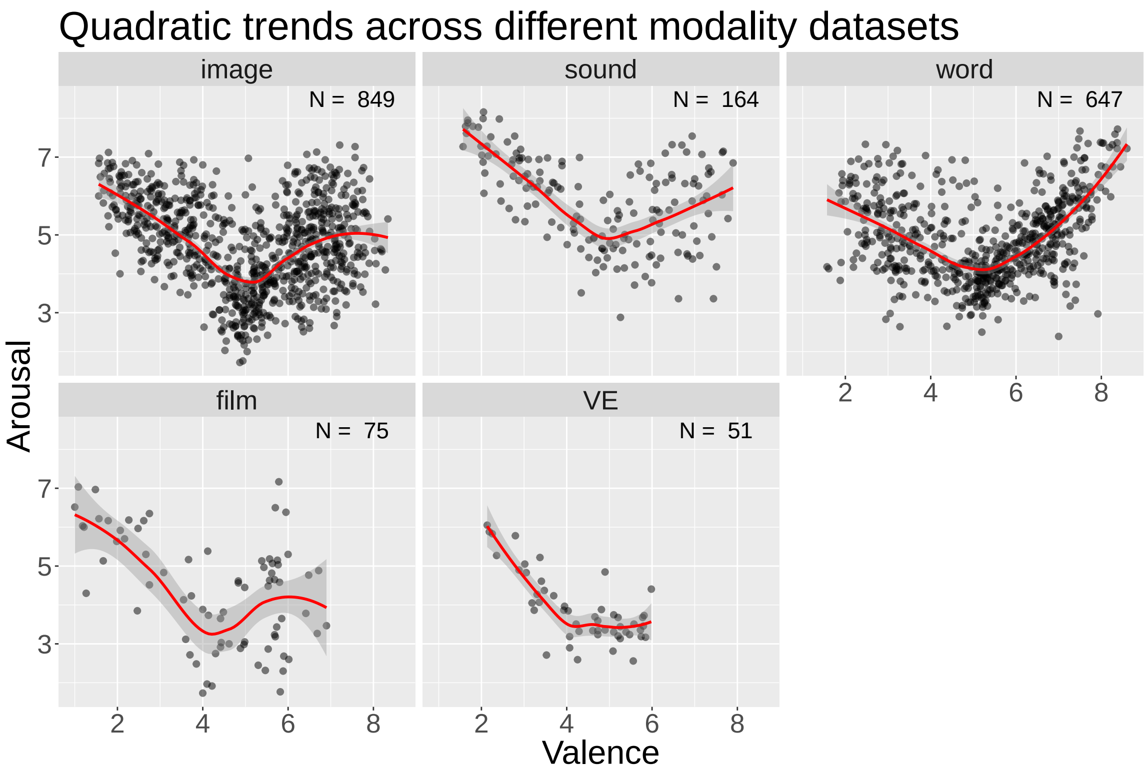


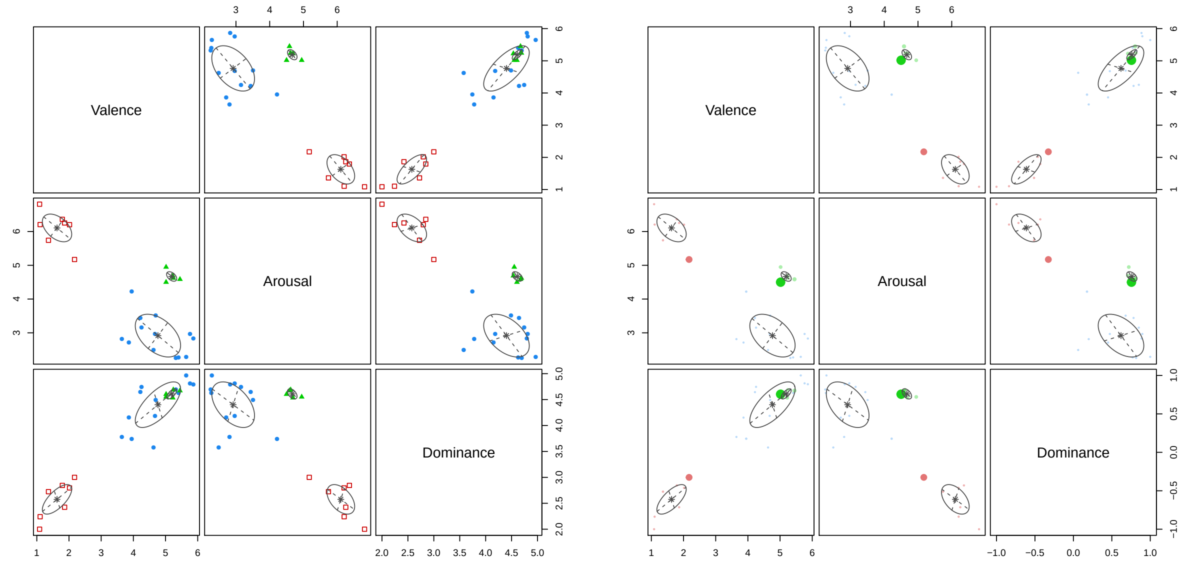
Figure 5.16: Five modality quadratic trends. Even if VEs are severely restricted in terms of both very high and very low Valence, they too present a quadratic relationship between Valence and Arousal, similarly to the other modalities.

Table 5.14: BIC values for MBC models, upon clustering exclusively films in this study. The top three best-fitting models are signalled with a bold font.

	EII	VII	EEI	VEI	EVI	VVI	EEE	EVE	VEE	VVE	EEV	VEV	EVV	VVV
1	-271.067	-271.067	-270.175	-270.175	-270.175	-270.175	-194.738	-194.738	-194.738	-194.738	-194.738	-194.738	-194.738	-194.738
2	-197.503	-194.007	-190.094	-189.049	-195.726	-194.198	-181.554	-187.27	-178.84	-183.591	-188.254	-182.94	-194.131	-188.187
3	-189.38	-169.624	-189.277	-170.179	-182.478	-180.425	-178.127	-186.239	**-156.751**	-169.653	-194.87	-169.781	-196.162	-187.172
4	-175.163	**-160.169**	-178.453	**-160.784**	-192.31	-175.499	-170.948	-172.809	-161.292	-165.517	-201.934	-179.645	-176.512	-195.672
5	-179.707	NA	-193.679	NA	-208.427	-197.423	-175.265	NA	NA	NA	-192.846	-182.235	NA	NA
6	-181.268	NA	-187.463	NA	NA	NA	-191.563	NA	NA	NA	-184.801	NA	NA	NA
7	-178.913	NA	-182.829	NA	NA	NA	-187.614	NA	NA	NA	-196.319	NA	NA	NA
8	-174.97	NA	-176.485	NA	NA	NA	-182.773	NA	NA	NA	-199.777	NA	NA	NA
9	-174.028	NA	-175.14	NA	NA	NA	-180.834	NA	NA	NA	-215.008	NA	NA	NA

Table 5.15: Film-only cluster centroids and sizes, for $k = 3$.

Dimension	Cluster 1 ($N = 13$)	Cluster 2 ($N = 7$)	Cluster 3 ($N = 5$)
Valence	4.764	1.628	5.190
Arousal	2.917	6.107	4.669
Dominance	4.402	2.576	4.610



(a) Film classification into 3 clusters.

(b) Stimulus membership uncertainty: the larger and darker the points, the higher the uncertainty.

Figure 5.17: Film clusters created using model-based clustering.

We also checked how closely these results reproduce those from Study 3A (see Section 4.3.4.1, p. 176). This was done by visualising Study 3A and Study 3B film data side by side, and in each case, coding the five *expected / intended* film clusters using different *symbols*, and the different *empirical / observed* clusters using different *colours*: 4 for Study 3A, and 3 for Study 3B (see Figure 5.18, p. 264).

5.3.8 Predicting engagement scores

Because it was not possible to directly match films to VEs according to the PAD model, any comparisons in terms of presence/engagement (or ITC scores) are more prone to error. For this reason, at the onset any models predicting either engagement or physical symptoms also included interaction terms between stimulus types (i.e., films and VEs) and the PAD dimensions. A further interaction was included between stimulus type and the complementary ITC factor score, e.g., if predicting Presence/Engagement scores, we tested the interaction between stimulus modalities and ITC Physical Effects - and vice versa. Other covariates were included as well, such as participant-level covariates (e.g., baseline Valence, Age, or PHQ-8 depression scores)[18]

All continuous predictors were standardised. The covariate for frequency of computer

[18] VE similarity was not included here as a covariate because it only presents values for VEs, but not for films which were constant throughout the study. However, this measure will indeed be used as a covariate in future models - see Section 7.3.2.2, p.302.

use (`CompFreq`) was not included in these models, because it has 0 variability (every participant gave a rating of 4 on this measure). Models were again computed using linear-mixed effects and `R` package `lme4`, with valuable predictors identified using the package `LMERConvenienceFunctions` (Tremblay & Ransijn, 2015). Based on a full model (illustrated in Listing 5.1, p. 265), backward-stepwise elimination was used to filter out unnecessary predictors, in such a way as to achieve improvements in AIC values. This was done separately for ITC engagement and physical symptoms outcomes.

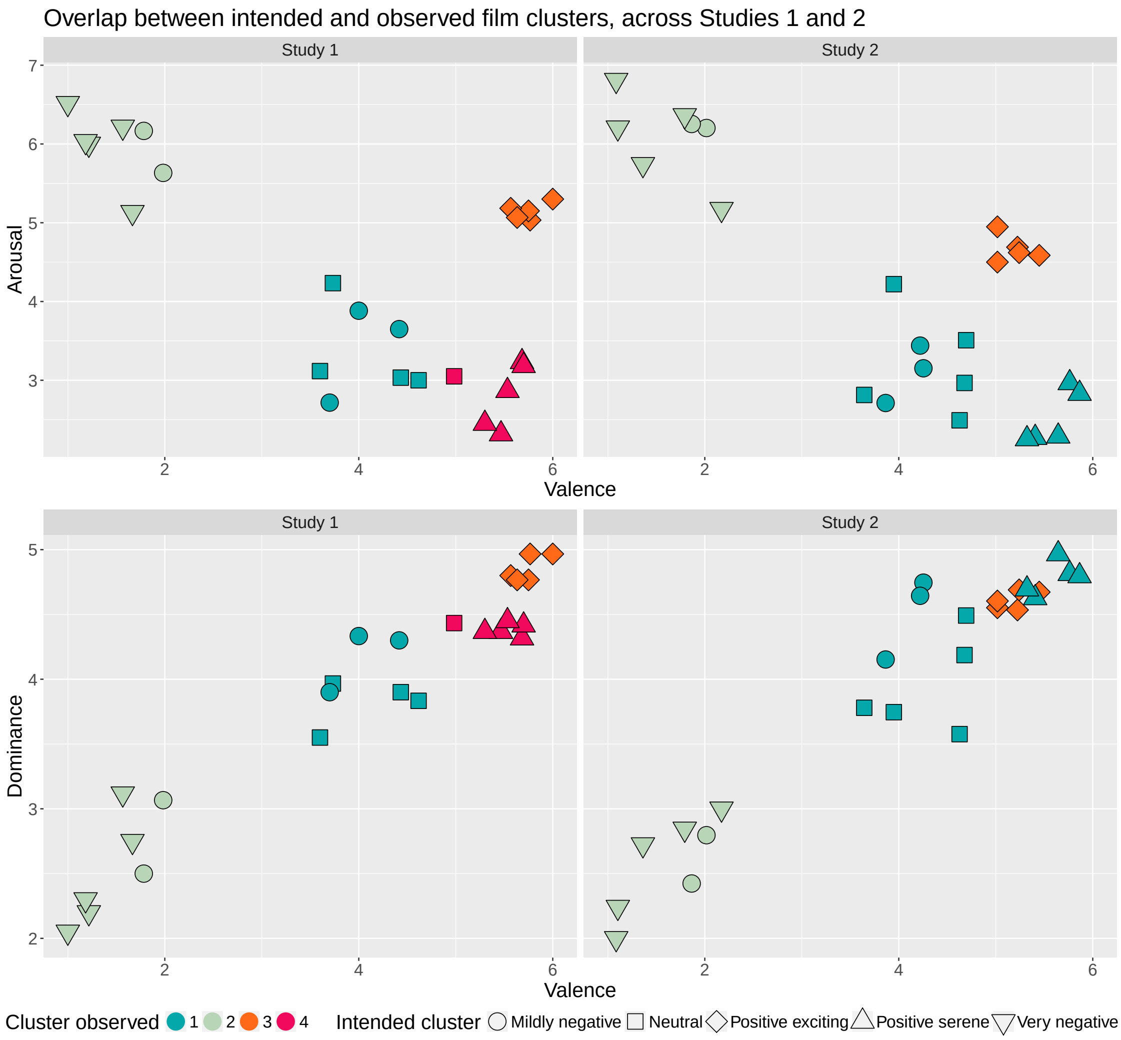


Figure 5.18: Overlap between intended and observed film clusters, shown separately for Study 3A and the current Study 3B. When points are grouped by colour, this shows their empirical / observed classification via MBC. When grouped by shape, this shows the expected classification of these stimuli, upon selecting them on YouTube or Second Life to match the other pre-existing modalities. In some cases, the two classifications can differ considerably: e.g., for Study 3A, the expected mildly negative cluster (in circles) is split between two other empirical clusters - the intensely negative and the neutral one (blue and grey). For Study 3B, the same is true, but in addition, the expected positive and serene cluster (triangles) has also merged with the expected neutral cluster (squares).

Listing 5.1: `R` code snippet: General form for the `lmer()` models used to predict ITC factor scores

```
full_phys_symp_model_51 <- lmer( ITC_Presence_Factor ~
                                    as.factor( stimulusClusterNumber ) +
                                    as.factor( stimulusBlockType ) * s( rateV ) +
                                    as.factor( stimulusBlockType ) * s( rateA ) +
                                    as.factor( stimulusBlockType ) * s( rateD ) +
                                    s( rateBaseV ) + s( rateBaseA ) + s( rateBaseD ) +
                                    as.factor( session ) +
                                    s( Age ) + as.factor( Gender ) + as.factor( Nationality ) +
                                    s( PhotoFreq ) + s( FilmFreq ) + s( CompComf ) +
                                    s( GameFreq ) + s( MMORPGFreq ) +
                                    s( VWFreq ) + s( SLFreq ) +
                                    as.factor( stimulusSoundFilmVRMusic ) +
                                    s( PANAS_Neg ) + s( PANAS_Pos ) + s( PHQ_total ) + s( TAS20_DDF ) + s( TAS20_DIF ) +
                                    s( ITC_PhysicalSymptoms_Factor ) * as.factor( stimulusBlockType ) +
                                    ( stimulusClusterNumber | subject_nr ) +
                                    ( 1 | New.SLURL ),
                                  data = data_for_ITC_models,
                                  verbose = TRUE, REML = FALSE,
                                  control = lmerControl( optimizer = "bobyqa",
                                                         optCtrl = list( maxfun = 50000 ) ) )
```

Table 5.16 below contains the best fitting models for each ITC outcome variable. Because stimuli could not be matched `1:1`, the models include data referring to all 51 VEs and 25 films clips. For this same reason, we also included interaction terms between the (unmatched) modalities and PAD scores.

Table 5.16: Predicting ITC engagement and physical symptoms for film and VE data, with Cluster 5 (neutral) used as a baseline, and films used as the baseline for stimulus types.

	Presence/Engagement model	Physical Symptoms model
(Intercept)	7.22^{***}	0.46^{***}
	(0.34)	(0.10)
Cluster 1 (mildly negative)	0.21	
	(0.17)	
Cluster 2 (intensely negative)	0.48^{**}	
	(0.18)	
Cluster 3 (positive and exciting)	-0.89^{***}	
	(0.19)	
Cluster 4 (positive and serene)	0.46^{**}	
	(0.16)	
Stimulus type: VE	-1.61^{***}	0.32^{***}
	(0.12)	(0.03)
Valence	0.36^{***}	
	(0.04)	
Arousal	0.24^{***}	
	(0.04)	
Baseline Valence	0.13^{*}	-0.11^{***}
	(0.05)	(0.02)
Stimulus music: present	-0.53^{***}	
	(0.15)	
Baseline Arousal		0.19^{***}
		(0.02)
Baseline Dominance		0.04^{*}
		(0.02)
Stimulus type: film × Valence		-0.15^{***}
		(0.02)
Stimulus type: VE × Valence		-0.04^{*}
		(0.02)
Stimulus type: film × Arousal		0.15^{***}
		(0.02)
Stimulus type: VE × Arousal		-0.04^{*}
		(0.02)
AIC	19009.42	10736.28
BIC	19092.43	10812.91
Log Likelihood	-9491.71	-5356.14
Num. obs.	4384	4384
Num. groups: Stimulus ID	76	76

Num. groups: Participant	59	59
Var: Stimulus ID (Intercept)	0.14	0.00
Var: Participant (Intercept)	5.48	0.55
Var: Residual	4.09	0.63

$^{***}p < 0.001$, $^{**}p < 0.01$, $^{*}p < 0.05$

5.4 Discussion

Despite originating from different datasets and participant samples, we can draw direct comparisons between image, word, sound, film and VE ratings, because the two samples were shown to be almost indistinguishable on a variety of covariate measures, including: age, computer / console gaming frequency, depression scores and alexithymia scores etc. The only exception to this was the significant difference we found between the samples for baseline Dominance, but only when participants returned to the study for their second session. Despite this, whenever all five modalities were compared, baseline Dominance was never an a-priori point of focus for our analysis, and so is assumed to have carried little to no weight. On the other hand, when predicting participant presence and physical symptoms scores, we *did* assess the impact of baseline Dominance (measured when participants began each study session) - although in this case, the data originated from the same sample, hence the between-sample differences are of no consequence.

Given this, it is interesting to mention that our 25 films were rated very similarly by the two samples - with correlations equal to, or over, 0.95 for all three PAD dimensions. This suggests that these film stimuli are highly reliable in terms of emotion elicitation, and also that the `optmatch` matching procedure from Study 3A (see Section 4.3.2, where this was used to select the 25 films out of the complete pool of 75) yielded useful, usable results.

In terms of further adding VEs as the fifth modality to these pre-existing sets of words, sounds, images, and films, we noted that this was not feasible due to several factors. Firstly, when comparing univariate PAD distributions, VEs occupied relatively more narrow portions of PAD space, compared to the other modalities. Especially in terms of Valence, VEs did not include stimuli as negative as some of our images or sounds, for example. Also in terms of high Valence scores, VEs again showed ranges which were more restricted relative to the other modalities. Arousal yielded a similar pattern, with VEs unable to elicit scores either as high, or as low as films, for example. Similar findings were found in the case of low Dominance as well - which images seemed much better suited to tap into.

Secondly, when comparing PAD bivariate relationships between the stimulus modalities, it becomes apparent that certain areas within 2D PAD space are void of VEs, yet include other types of stimuli, for instance: very low Valence and very high Arousal; very high Valence and very high Arousal; high Dominance and average Arousal; low Dominance and average Arousal; high Arousal and low Dominance; and finally, very low Valence and very low Dominance. Relatedly, the shape of these relationships is also different across modalities: interestingly, VEs form a less typical pattern of negative linear correlation between Valence and Arousal, but only so long as Valence remains below the scale midpoint (i.e., for scores between 1 and 5). Afterwards, for higher values of Valence, the correlation levels off to show no particular trend (i.e., Arousal then remains fairly constant, despite the increase in Valence). This forms a contrast with the other modalities which all show (to varying degrees) a "U"-shaped, quadratic trend between Valence and Arousal.

For these reasons, two courses of action were determined: firstly, that `1:1` matching of VEs to films (or any other stimuli) was not likely to produce useful results - hence it was replaced by further exploring VEs via MBC instead; and secondly, that a future study should investigate if wearing an HMD (head-mounted display) can correct these range restriction issues, and induce higher variance / more extreme ratings for VEs.

Upon clustering the VEs, only two components were discovered in the data - which contrasts sharply with the $k = 4$ solution previously (and repeatedly) found for other modalities. Multiple reasons may exist for this: perhaps VEs are stereotypically associated with game-like experiences, and specific types of content / activities, which could explain why, e.g., PAD areas with average Arousal and higher Dominance are not populated by VEs. Also, because games and game-like environments often do not emphasise character personas (to make them more human) or emotional situations (Stuart, 2016), participants may come to expect this and focus more on tasks to achieve, enemies to defeat etc, rather than undergoing actual emotional experiences. It is this shift in focus, and this form of "emotional blunting" which might be underlying these ratings.

On the other hand, it could also be the case that pre-existing Second Life environments might simply not be as suitable to elicit emotional experiences as custom-made VEs. This is because, even if admitting that some form of blunting occurs for the Valence dimension, one might still expect that high Arousal zones would be more populated even if stimuli are likened to games (e.g., due to any thrilling experiences of fear, or excitement after having successfully addressed various 'tasks' in a game etc). Despite this, these extremities for the Arousal dimension remain almost as poorly populated as for Valence. Finally, these VEs may also show more variation on affect dimensions other than PAD, which were not measured here (such as 'novelty' as inspired by some appraisal theories, Moors, 2009).

Other deviations from expected results were also found in the case of films - albeit far less concerning ones. Despite expecting a four-cluster solution for films based on our previous results, only three clusters were suggested this time. This is because two expected clusters were close enough together to be merged into a single cluster: the neutral and the positive serene clusters. Any implications for this discrepancy are moderated by the fact that next best-fitting models for this data were indeed $k = 4$ solutions, and by the fact that the expected and unmerged clusters are still visible as separate entities, e.g., in the Valence × Arousal panel from Figure 5.17 (p. 262). Hence, this $k = 3$ solution can be seen as a related and less likely alternative to previous $k = 4$ solutions, rather than an opposing result.

Finally, in terms of presence and physical symptoms scores, there was fairly little overlap in terms of the predictors / covariates proven to be useful for both models simultaneously. Only stimulus modality and baseline Valence were significant predictors in both models. Overall, VEs were more prone to inducing negative physical symptoms than films, and (surprisingly), were also *less* engaging than films - thus contradicting our original expectation. Higher baseline Valence upon starting the study sessions was also significantly associated with feelings of higher presence, as well as fewer physical symptoms.

For specific presence predictors, it was shown that the expected clusters - while containing some error since they were not fully replicated empirically - were still useful in the model, probably since they encapsulate simultaneous variations in all three PAD dimensions. As such, both negative clusters led to higher presence compared to the neutral cluster (although only the intensely negative one - significantly so), whereas the two positive clusters showed a divergent pattern, with the positive and exciting cluster being less engaging, and the positive and serene one - more so, than the baseline. This pattern is probably due to the fact that the positive and exciting cluster included various erotic stimuli, which might not be as believable or as comfortable to watch in a formal university setting. For this engagement model, Valence and Arousal also generally (and independently) contributed to the prediction, in that the more positive and alerting the stimulus, the more engaging it was. Interestingly, for stimuli where background music was included, these were generally *less* engaging than those without music. It is difficult to interpret the reasons behind this, however it could be due to the music perhaps not being a good match for the content of the stimulus, or being seen as distracting.

For Physical Symptom-specific predictors, higher baseline Arousal and Dominance were associated with more intense negative effects - and while interesting, this is a fairly unclear and unexpected. In addition, several interactions were found here between stimulus modality and Valence / Arousal: for films, more positive Valence led to fewer physical symptoms, whereas higher Arousal had the opposite influence. For VEs on the other hand, both higher Valence and higher Arousal were slight protective factors against developing physical symptoms in VR.

Chapter 6

Study 3C: Using immersive VEs as affective stimuli

6.1 Introduction

NAVIGATING an 'emotion-inducing' VE displayed on an external computer monitor should already present desirable properties such as allowing participants to develop a sense of agency - where they alone decide what actions to take within the VE and how to explore it. Despite being present in real life, this property is lacking in multiple other forms of emotion elicitation (e.g., images or films), but VEs embody it by definition. Under these circumstances, it is unclear whether presenting VEs using a Head-Mounted Display (HMD) instead of a computer monitor can offer any *additional* benefits - particularly considering the risk that motion sickness can occur with prolonged use.

However, due to the findings from Study 3B above, we would like to consider the possibility that using an HMD may 'intensify' emotional ratings of participants immersed in virtual worlds. Thus, we designed a new study in order to address whether wearing an HMD can induce higher variance / more extreme ratings for VEs, compared to a computer screen, and whether this gain was enough to justify using an HMD, despite the risks of negative physical symptoms (i.e., simulator sickness) appearing and potentially distorting ratings (of course, above and beyond any ethical concerns).

6.1.1 Aims

In order to investigate the impact of wearing an HMD vs. using an external monitor on PAD ratings, as well as on presence and negative symptoms, in this study we collected data on a sub-selection of VEs used previously in Study 3B. In reusing certain VEs with an HMD, our aim was to draw comparisons between the two conditions. In this same study (hither forth referred to as Study 3C), we also assessed which measures (including participant-level covariates such as age, depression scores, etc.) can predict our outcomes across both samples (i.e., Valence, Arousal, Dominance, engagement, and physical symptoms ratings).

6.2 Method

6.2.1 Participants

Via an advertisement posted to the University's Careers Service, we recruited a sample of $N = 60$ participants for this study (compensated £7/h, according to the level of minimum wage). One participant did not respect the study instructions, and did not use the Head-Mounted Display to navigate VEs. Because of this, their data were excluded and one additional participant was recruited. The final sample is described in Tables 6.1 and 6.2 below, in terms of their responses to the battery of questionnaires used consistently throughout this work:

Table 6.1: Study 3C sample description and background questionnaire measures.

No	Measure	Mean	Trim	Median	SD	Min	Max	Range	Skew	Kurt	SE
1	Age	23.67	23.04	22.00	4.43	18.00	41.00	23.00	1.51	2.64	0.57
2	CompComf	3.80	3.92	4.00	0.48	2.00	4.00	2.00	-2.31	4.64	0.06
3	CompFreq	3.97	4.00	4.00	0.18	3.00	4.00	1.00	-5.07	24.11	0.02
4	FilmFreq	2.57	2.58	3.00	0.72	1.00	4.00	3.00	-0.23	-0.28	0.09
5	GameFreq	1.93	1.92	2.00	1.35	0.00	4.00	4.00	0.12	-1.25	0.17
6	MMORPG Freq	0.58	0.44	0.00	0.81	0.00	3.00	3.00	1.25	0.78	0.10
7	PhotoFreq	2.77	2.85	3.00	0.91	0.00	4.00	4.00	-0.74	0.35	0.12
8	SLFreq	0.10	0.00	0.00	0.35	0.00	2.00	2.00	3.65	13.59	0.05
9	VWFreq	0.62	0.46	0.00	0.83	0.00	4.00	4.00	1.68	3.58	0.11
10	PANAS_Neg	6.03	5.83	5.50	3.08	1.00	14.00	13.00	0.61	-0.08	0.40
11	PANAS_Pos	12.72	12.88	13.00	3.25	5.00	19.00	14.00	-0.42	-0.47	0.42
12	PHQ_total	5.57	4.92	4.00	4.48	0.00	17.00	17.00	1.10	0.40	0.58
13	TAS20_DIF	7.45	7.12	6.50	5.45	0.00	21.00	21.00	0.44	-0.70	0.70
14	TAS20_DDF	7.77	7.67	7.00	2.68	3.00	13.00	10.00	0.33	-0.95	0.35
15	Base V	5.42	5.42	5.00	1.08	2.00	8.00	6.00	-0.23	0.97	0.14
16	Base A	3.53	3.44	4.00	1.89	1.00	8.00	7.00	0.26	-0.94	0.24
17	Base D	4.88	4.88	5.00	1.64	1.00	8.00	7.00	-0.04	-0.11	0.21

Note. **Base A** = Baseline Arousal; **Base D** = Baseline Dominance; **Base V** = Baseline Valence; **CompComf** = level of comfort when using a computer; **CompFreq** = frequency of using a computer; **FilmFreq** = frequency of watching films; **GameFreq** = frequency of gaming; **MMORPGFreq** = Massively Multiplayer Online Role-Playing Games gaming frequency; **PANAS_Neg** = PANAS Negative Affect scale; **PANAS_Pos** = PANAS Positive Affect scale;

PhotoFreq = frequency of taking photos; **PHQ_total** = PHQ-8 total score; **SLFreq** = frequency of using Second Life; **TAS20_DIF** = TAS-20 Difficulty Describing Feelings subscale; **TAS20_DIF** = TAS-20 Difficulty Identifying Feeling subscale; **VWFreq** = frequency of using virtual worlds.

Table 6.2: Study 3C sample description, in terms of gender and nationality.

Nationality	Female	Male
Africa	1	1
Australia	2	1
Eastern Europe	3	2
Far East	5	5
North America	4	4
South America	0	1
Western Europe	15	16

6.2.2 Stimuli and apparatus

This study used the Oculus Rift DK1 Head-Mounted Display (HMD) in order to generate an immersive experience for participants. The Oculus Display Driver (v. `1.2.8.0`), the Oculus Rift Sensor Driver (v. `1.0.14.0`), and Oculus Runtime (v. `5.0.1` and following a bug, v. `0.8.0`) were used alongside this hardware. The HMD was paired with the set of lenses fitted by default, and a calibration process was carried out pre - data collection.

In addition, this HMD presents a screen resolution of 1280 × 800, 7" screen size, latency of 50-60ms, a field of view (FOV) of 110°, and, of course, stereoscopic visual projection. The DK1 is fitted with a gyroscope, accelerometer, and magnetometer, and is capable of orientation tracking, although not positional tracking. This means that participants can change their direction of gaze anywhere around them (by rotating their head to look upwards, downwards, or sideways), and the visual feedback on the Rift will change accordingly. However, changing the position of the head itself (e.g., when bending their whole body forward, or leaning to the side) will not be tracked by the Rift and hence, will not adjust visual input in any way. The latter issue, jointly with the latency, can contribute to simulator sickness.

For this study we used 6 stimuli, i.e., 3 VEs which presented the lowest uncertainty within the general positive cluster, and another 3 - within general negative cluster discovered previously via MBC (see Section 5.3.5, and Table 5.11, p. 256). For the positive cluster, the 3 representative VEs are: **Kalepa** (a Mediterranean landscape with bright sun, lush vegetation, flowers, and a stone bridge extending over a body of water), **Yumix Prada** (a replica of Venice, with gondolas, romantic-looking buildings, and restaurants facing the water), and **Brandy Wine Island** (a rocky beach with waves washing up on the shoreline, a lighthouse, as well as a small campfire and some sun beds with cushions).

In terms of the negative cluster, the 3 most representative VEs sampled for use in this study were: **Twilight Hollow** (a horror world where monsters chase after users, and strange occurrences appear in haunted spaces), **Picklemoon** (a haunted mansion with chilling background creaks and groans, fiends, and anatomical curiosities on display), and **Hell** (a dark cave / inside of volcano with corpses hung up for torture). Because at the time of data collection, **Twilight Hollow** had been taken offline, we replaced it with the next VE in line in terms of uncertainty: **PickleSong** (an eerie autumn forest, including a haunted mansion where ghostly presences float mid-air, hooded figures perform rituals, and dead bodies are scattered around). Illustrative scenes for each of these VEs are included in Appendix E.3 (p. 481).

In order to display these VEs and output their binocular projection to the HMD, we used a Second Life viewer program designed for Windows, named CtrlAltStudio (`v1.2.5.43397 Alpha`). In parallel, our setup included a desktop screen recorder (`Movavi`). This was used to track participants' activity through the VEs and to check whether they were conforming to the study tasks as instructed, and whether they were having difficulties in using the software. Equally, this footage was used as a way to account for any unusual values (i.e., by checking what participants actually did in the associated VE), or whether they respected the time limit imposed for exploring it etc. Participants themselves were never filmed (no "camera" was actually active in the room). Finally, to record responses and control the random order of presentation of the VEs themselves, we used OpenSesame, an open-source Python-based experiment builder. While presenting a GUI, OpenSesame also provides the option for directly embedding custom Python code. In this study, this feature was essential in order to: switch window focus from OpenSesame to CtrlAltStudio, send the required keyboard shortcut to switch CtrlAltStudio from normal viewing mode into Rift mode, and finally, when navigation time elapsed for each VE, custom Python code as also used to return CtrlAltStudio to normal viewing mode, and switch focus back to the OpenSesame window for collecting ratings.

6.2.3 Instruments

The same measures were collected as in the previous two studies. For a discussion, please refer back to Study 3A's Section 4.2.3 (p. 157).

6.2.4 Procedure

Participants were greeted by the researcher and were provided with a standard set of instructions included in Appendix E.2, and answered any questions. Subsequently, participants were given a VR orientation/practice phase consisting of only 1 VE, which is a Second Life representation of the University of Edinburgh[1]. Participants were allowed to explore this VE using the Oculus Rift for 1.5 minutes, after which an alarm signalled that the time had elapsed. However, in cases where using the time proved insufficient for the participant to gain an adequate understanding for how to control their avatar, use the Rift, or submit ratings, this practice was extended as necessary.

Once the practice phase was complete, the participants continued with the real experiment which opened with a battery of psychometric measures (all of which have been used and described in previous chapters of this thesis). As discussed, the study was implemented on a Windows system, with a suite of programs used jointly to perform various functions: Movavi desktop recorder, CtrlAltStudio Second Life Viewer, and OpenSesame (as the experiment builder).

More specifically, participants would finish their questionnaires in OpenSesame and then move on to exploring, and then rating, VEs. At the beginning of each VE trial, a button was presented to them in OpenSesame, and once pressed, it would switch to Rift mode within the separate Second Life viewer, and then pause for 10s before actually unveiling the online VE to participants in order to give the objects within a chance to load fully. The pressing of the button would have two further consequences once the 10s had elapsed: switching window focus[2] from OpenSesame to the Second Life viewer, and starting a parallel countdown within OpenSesame for the 1.5 minutes afforded to explore each VE.

Once this time passed, a different sequence of processes would occur: an alarm would ring for 2s (which is identical to Study 3B), and a Python script within OpenSesame would again shift window focus, this time maximising a rating screen on the external computer monitor. In parallel, CtrlAltStudio would drop out of Rift mode and the participant's avatar would be teleported back to a "baseline" VE, where no sound was present (i.e., Edinburgh University, which was also used in the practice). Abandoning Rift mode between trials also had the benefit of resetting any unusual angles from the past trial.

Because this study included only 6 VEs, the stimuli were randomised individually rather than by (and within) blocks. This is because with such a short testing time, it is unlikely that any emotional "fatigue" would have a chance to set in and affect ratings. To minimise emotional carry-over between stimuli, we interspersed the exploration of any two VEs with a 3-minute buffer task. This had the additional benefit of reducing the chances of simulator (or motion) sickness appearing with prolonged use of the Rift. Also, a specific "break task" ensured that the testing conditions remained constant across

[1] Found at this URL within Second Life: http://maps.secondlife.com/secondlife/Edinburgh%20East/135/117/26

[2] Using a Python inline script involving packages `win32gui`, `win32con`, `re`, `win32com.client`, and `time`.

participants. For this purpose, an unrelated colouring task was used: participants were given coloured pencils and sheets with a variety of abstract geometric shapes, and asked to spend their 3-minute break colouring them in, in any way they saw fit. This was deemed as unlikely to affect emotional ratings, and had no relation to any of the stimuli or experimental tasks.

Finally, the same Second Life settings as in Study 3B were used here: participant avatars were matched to their real genders, navigation speed was set to "running", and the time of day was fixed to midday.

6.3 Results

6.3.1 Sample comparisons

In order to safely compare participant samples from the current and previous Study 3B, we investigated if there were any differences between these samples, on a variety of measures. Results from Table 6.3 suggest that there are no significant differences, hence data can safely be pooled across studies, without the need to incorporate the associated interactions in later models. This will allow a direct comparison of data collected on VEs displayed on a large monitor (i.e., Study 3B), with VEs displayed on an HMD, i.e., the Oculus Rift DK1 (i.e., current study).

Table 6.3: Comparing continuous, single-measure, covariates in Study 3B vs Study 3C. The table represents *t*-test results when checking for differences between the two samples.

No	Measure	Δ	μ_1	μ_2	t	p	df	CI lower	CI upper
1	Age	1.09	23.67	22.58	1.37	0.17	116.92	-0.49	2.67
2	CompComf	0.09	3.80	3.71	1.03	0.31	116.87	-0.08	0.26
3	CompFreq	-0.03	3.97	4.00	-1.43	0.16	59.00	-0.08	0.01
4	FilmFreq	0.02	2.57	2.54	0.18	0.86	116.16	-0.25	0.30
5	GameFreq	0.37	1.93	1.56	1.59	0.11	115.95	-0.09	0.84
6	MMORPGFreq	-0.06	0.58	0.64	-0.36	0.72	111.51	-0.39	0.27
7	PhotoFreq	0.02	2.77	2.75	0.11	0.91	112.57	-0.34	0.39
8	SLFreq	0.03	0.10	0.07	0.52	0.60	115.79	-0.09	0.15
9	VWFreq	0.07	0.62	0.54	0.49	0.62	117.00	-0.22	0.37
10	PANAS_Neg	-0.00	6.03	6.03	-0.00	1.00	117.00	-1.11	1.11
11	PANAS_Pos	-0.05	12.72	12.76	-0.07	0.94	112.32	-1.36	1.27
12	PHQ_total	-0.06	5.57	5.63	-0.08	0.93	110.88	-1.52	1.39
13	TAS20_DIF	0.33	7.45	7.12	0.33	0.74	116.90	-1.66	2.32
14	TAS20_DDF	0.17	7.77	7.59	0.35	0.73	116.94	-0.80	1.15
15	HMD vs Large monitor, session 1: Base V	0.15	5.42	5.27	0.65	0.52	111.20	-0.30	0.59

16	HMD vs Large monitor, session 1: Base A	-0.11	3.53	3.64	-0.34	0.74	116.07	-0.76	0.54
17	HMD vs Large monitor, session 1: Base D	0.43	4.88	4.46	1.50	0.14	115.81	-0.14	0.99
18	HMD vs Large monitor, session 2: Base V	0.18	5.42	5.24	0.95	0.35	116.43	-0.20	0.55
19	HMD vs Large monitor, session 2: Base A	0.38	3.53	3.15	1.22	0.23	111.87	-0.24	1.00
20	HMD vs Large monitor, session 2: Base D	0.32	4.88	4.56	1.17	0.25	114.35	-0.23	0.87

Note. **Base A** = Baseline Arousal; **Base D** = Baseline Dominance; **Base V** = Baseline Valence; **CompComf** = level of comfort when using a computer; **CompFreq** = frequency of using a computer; **FilmFreq** = frequency of watching films; **GameFreq** = frequency of gaming; **MMORPGFreq** = Massively Multiplayer Online Role-Playing Games gaming frequency; **PANAS_Neg** = PANAS Negative Affect scale; **PANAS_Pos** = PANAS Positive Affect scale; **PhotoFreq** = frequency of taking photos; **PHQ_total** = PHQ-8 total score; **SLFreq** = frequency of using Second Life; **TAS20_DIF** = TAS-20 Difficulty Describing Feelings subscale; **TAS20_DIF** = TAS-20 Difficulty Identifying Feeling subscale; **VWFreq** = frequency of using virtual worlds.

6.3.2 Comparing PAD and ITC scores between immersive and non-immersive VEs

We first conducted a visual examination of the pooled data in Figure 6.1 (p. 278), which suggests that, when VEs were described using Valence and Dominance and the Oculus Rift was used simultaneously, there appears to be better discrimination between VEs, relative to when a large computer screen was used instead (i.e., points are more spread out). From the same figure, it also appears that presence was higher when the Oculus Rift was used, especially for positive VEs.

Separately, we also investigated the pooled data across Studies 3B and 3C in terms of the average Valence, Arousal, Dominance, engagement and physical symptoms ratings received by each VE, and split by condition (i.e., large monitor vs. the Oculus Rift HMD) - see Figure 6.2 (p. 279). In this figure, the positive and negative VEs appear to perform as expected across the PAD model: regardless of whether or not an HMD was used, the positive and serene VEs score above average on the Valence and Dominance scales, but lower on the Arousal scale. In terms of the intensely negative VEs, these too perform as expected (again, regardless of display output): under average scores for Valence and Dominance, but above average on Arousal.

In the same figure, in terms of presence and physical symptoms, the data appear to be unsurprising: HMD wearers experienced more presence/engagement, but also more negative physical symptoms, compared to large screen users. However, the magnitude of differences between the HMD and large monitor output was not as obvious as may have been expected, and seemed to vary by VE: e.g., for **Yumix Prada** (the VE inspired by Venice), there was quite a marked difference in favour of the HMD (which appears to have intensified the ratings already received when using a large monitor), whereas for **PickleSong** (haunted mansion in a forest), the HMD seemed to have been somewhat less effective than the large computer screen.

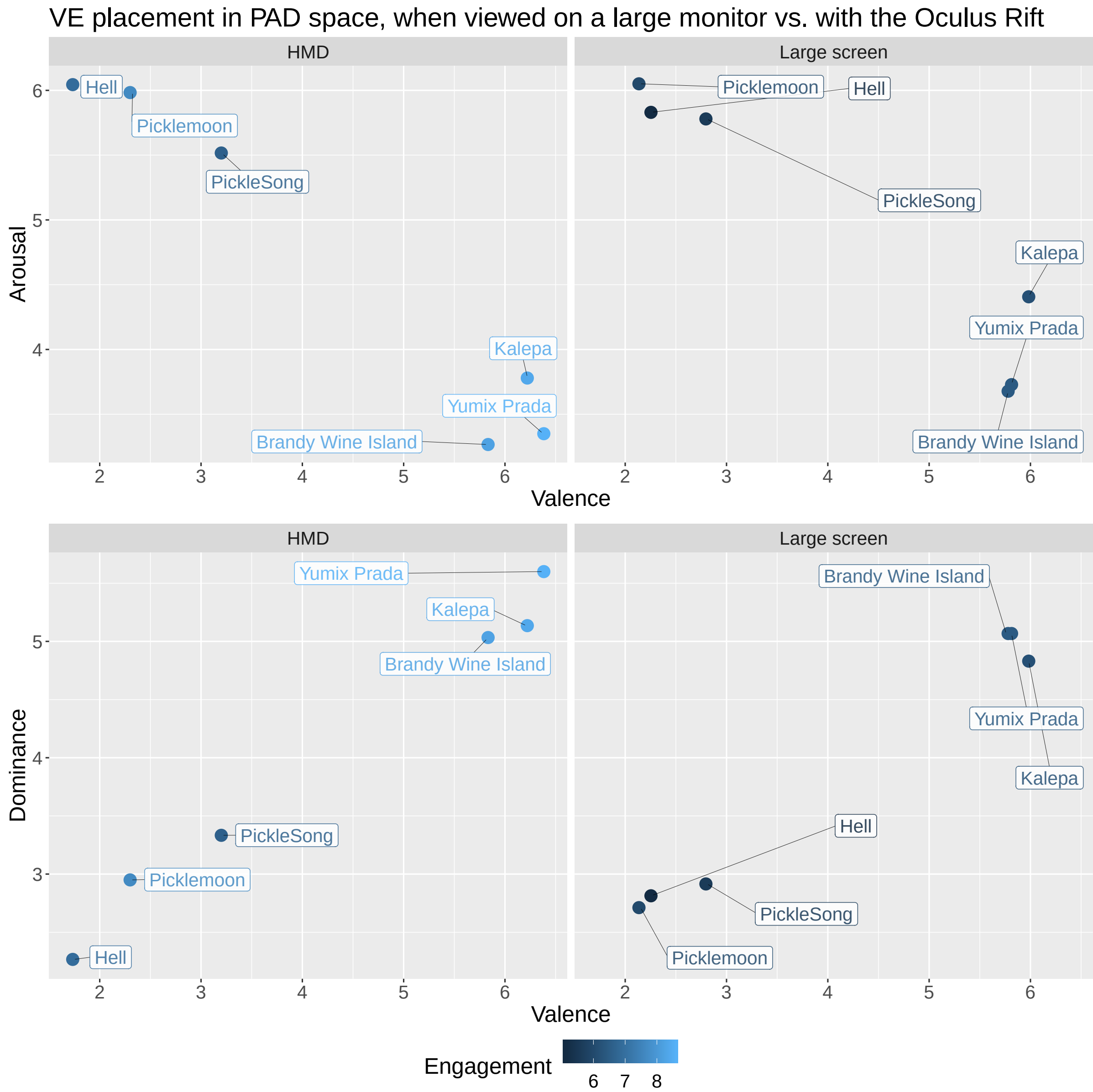


Figure 6.1: VE placement in PAD space, with or without an HMD for exploring the VEs. Individual VEs were also colour-coded based on the average level of presence/engagement they induced (as measured using the ITC-SOPI-SF items).

6.3.3 Predicting PAD scores

We checked whether any covariates / predictors were particularly useful for predicting Valence, Arousal and Dominance ratings received by the 6 VEs between conditions (with / without an HMD). The general format for model specification is show in Listing 6.1

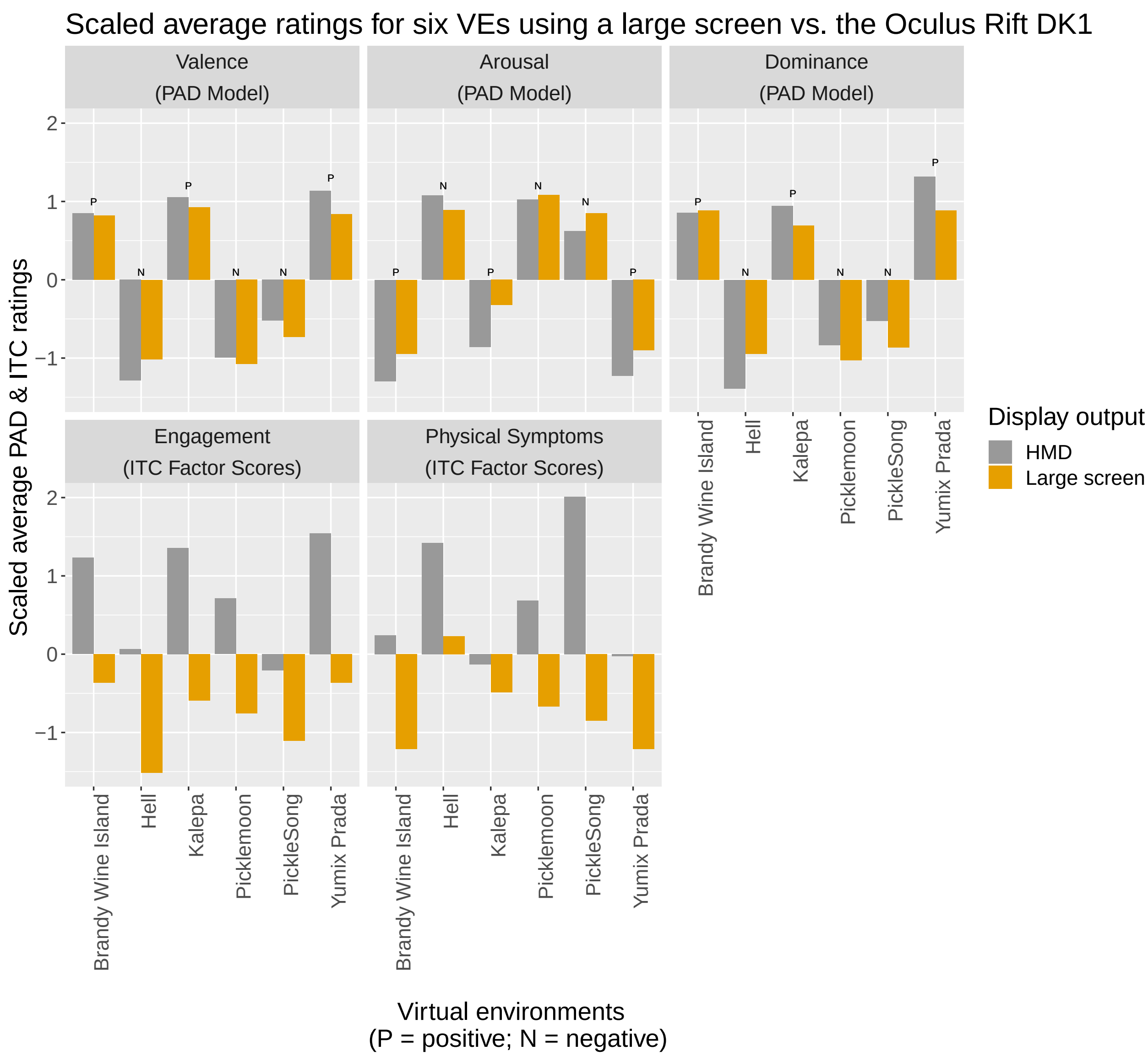


Figure 6.2: Scaled average stimulus ratings for VEs with and without an HMD. Bars extending under 0 represent under-average VE scores on the various measures.

(p. 280), and results are shown in Table 6.4 (p. 281).

Listing 6.1: `R` code snippet: General form for the `lmer()` models used to predict either PAD ratings, or ITC factor scores. Such full models were subjected to a backwards, stepwise selection algorithm within the `LMERConvenienceFunctions R` package. Within these models, all continuous predictors were scaled, and REML was switched to 'False' in order to be able to compare models based on AIC fit criteria. Finally, depending on the outcome variable assessed, the set of predictors will have been constructed to include interactions for the other remaining, and related, variables. Concerning the PAD model, for instance, Valence scores were predicted from, among other things, Arousal and Dominance-based interactions. Finally, while Nationality is included in the Valence model, it does not present any significant coefficients and can therefore be ignored.

```
full_Valence_model <- lmer( rateV ~
                                s(rateBaseV) + s(rateBaseA) + s(rateBaseD) +
                                s(Age) + as.factor(Gender) +
                                    as.factor(Nationality) +
                                s(PhotoFreq) + s(FilmFreq) + s(CompComf) +
                                s(CompFreq) + s(GameFreq) + s(MMORPGFreq) +
                                s(VWFreq) + s(SLFreq) +
                                s(PANAS_Neg) + s(PANAS_Pos) + s(PHQ_total) +
                                    s(TAS20_DDF) + s(TAS20_DIF) +
                                # Source referes to either: HMD or large monitor
                                as.factor( stimulusClusterNumber ) *
                                    as.factor(Source) +
                                s(ITC_PhysicalSymptoms_Factor) *
                                    as.factor(Source) +
                                s(ITC_Presence_Factor) * as.factor(Source) +
                                s(rateA) * as.factor(Source) +
                                s(rateD) * as.factor(Source) +
                                as.factor(stimulusSoundFilmVRMusic) +
                                # Effect of Source varies by VE.
                                ( 1 | New.SLURL ) +
                                ( 1 | Participant ),
                              data = combined_data,
                              verbose = TRUE, REML = FALSE,
                              control = lmerControl( optimizer = "bobyqa",
                                                      optCtrl = list( maxfun =
                                                          50000 ) ) )
```

Table 6.4: PAD models predicting Valence, Arousal and Dominance ratings received by the 6 VEs between conditions (with / without an HMD).

Coefficient	Valence model	Arousal model	Dominance model
(Intercept)	3.27^{***}	5.22^{***}	4.02^{***}
	(0.39)	(0.19)	(0.08)
Baseline Dominance	0.03		0.17^{*}
	(0.06)		(0.07)
Nationality: Australia	−0.20		
	(0.45)		
Nationality: Eastern Europe	0.29		
	(0.40)		
Nationality: Far East	−0.61		
	(0.38)		
Nationality: Middle East	−0.27		
	(0.50)		
Nationality: North America	0.05		
	(0.39)		
Nationality: South America	0.36		
	(0.56)		
Nationality: Western Europe	0.10		
	(0.37)		
Positive cluster	1.98^{***}	-0.92^{***}	
	(0.25)	(0.27)	
ITC Presence/Engagement Factor	0.21^{***}		
	(0.06)		
Arousal	-0.37^{***}		
	(0.06)		
Dominance	1.16^{***}		
	(0.06)		
Baseline Arousal		0.46^{***}	
		(0.09)	
Valence		-0.85^{***}	1.59^{***}
		(0.10)	(0.06)
AIC	2405.17	2768.71	2498.24
BIC	2477.78	2800.47	2525.47
Log Likelihood	-1186.58	-1377.35	-1243.12
Num. obs.	691	691	691
Num. groups: Participant	119	119	119
Num. groups: Stimulus ID	6	6	6
Var: Participant (Intercept)	0.07	0.63	0.41
Var: Stimulus ID (Intercept)	0.07	0.06	0.00
Var: Residual	1.69	2.67	1.83

$^{***}p < 0.001$, $^{**}p < 0.01$, $^{*}p < 0.05$

6.3.4 Predicting ITC scores

Using the `lme4` code demonstrated in Listing 6.2, we investigated which of a series of covariates and predictors (e.g., participant age, PAD ratings, display method: HMD vs. large monitor) predicted how presence- or physical-symptom inducing, the 6 VEs were. The best fitting model for each outcome is presented in Table 6.5 (p. 283). Interactions were incorporated into these models in order to check for any significant patterns suggested in Figure 6.2 (p. 279).

Listing 6.2: `R` code snippet: General form for the `lmer()` models used to predict ITC factor scores

```
full_immersion_model <- lmer( ITC_Presence_Factor ~
                                s(rateBaseV) + s(rateBaseA) + s(rateBaseD) +
                                s(Age) + as.factor(Gender) +
                                    as.factor(Nationality) +
                                s(PhotoFreq) + s(FilmFreq) + s(CompComf) +
                                s(CompFreq) + s(GameFreq) + s(MMORPGFreq) +
                                s(VWFreq) + s(SLFreq) +
                                s(PANAS_Neg) + s(PANAS_Pos) + s(PHQ_total) +
                                    s(TAS20_DDF) + s(TAS20_DIF) +
                                as.factor( stimulusClusterNumber ) *
                                    as.factor(Source) +
                                s(ITC_PhysicalSymptoms_Factor) *
                                    as.factor(Source) +
                                s(rateV) * as.factor(Source) +
                                s(rateA) * as.factor(Source) +
                                s(rateD) * as.factor(Source) +
                                as.factor(stimulusSoundFilmVRMusic) +
                                # Effect of Source varies by VE.
                                ( as.factor(Source) | New.SLURL ) +
                                (1 | Participant),
                              data = combined_data,
                              verbose = TRUE, REML = FALSE,
                              control = lmerControl( optimizer = "bobyqa",
                                                     optCtrl = list( maxfun =
                                                         50000 ) ) )
```

Table 6.5: Presence and physical symptoms models for Study 3B vs. Study 3C.

	Presence / Engagement Model	Physical Symptoms Model
(Intercept)	7.66^{***}	0.82^{***}
	(0.41)	(0.09)
Baseline Dominance	0.22	
	(0.16)	
Source: Large monitor	-1.65^{***}	
	(0.47)	
Valence	0.45^{***}	-0.09^{**}
	(0.11)	(0.03)
Arousal	0.28^{**}	
	(0.10)	
Baseline Valence		-0.18^{**}
		(0.06)
Baseline Arousal		0.21^{**}
		(0.07)
ITC Presence/Engagement Factor		-0.12^{**}
		(0.04)
AIC	3064.80	1749.75
BIC	3110.19	1786.05
Log Likelihood	-1522.40	-866.87
Num. obs.	691	691
Num. groups: Participant	119	119
Num. groups: Stimulus ID	6	6
Var: Participant (Intercept)	5.40	0.98
Var: Stimulus ID (Intercept)	0.42	0.00
Var: Stimulus ID × Source: Large monitor	0.11	
Cov: Stimulus ID (Intercept) × Source: Large monitor	-0.22	
Var: Residual	3.11	0.45

$^{***}p < 0.001$, $^{**}p < 0.01$, $^{*}p < 0.05$

6.4 Discussion

Firstly, in this study we confirmed that there were no significant differences between the Study 3B and Study 3C samples on any of the continuous covariates recorded (e.g., age or depression scores etc.). Because nationality restrictions were not imposed on participants, this feature therefore varied between the two studies. However, this did not influence results given that no significant main effects were discovered in relation to nationality in subsequent models where PAD dimensions / ITC factor scores were predicted.

Regardless of display output (HMD vs. large monitor), the two clusters performed as expected across the PAD model: the positive and serene VEs scored above average on the Valence and Dominance scales, but lower on the Arousal scale. In terms of the intensely negative VEs, these too followed an expected pattern: under average scores for Valence and Dominance, but above average on Arousal. In terms of presence and physical symptoms, results were also unsurprising: HMD-wearers experienced more presence, but also more negative physical symptoms, compared to large screen viewers.

Despite some trends appearing within the data, the difference between the HMD and large monitor output could not be detected in mixed-effects models: the impact of wearing an HMD vs. navigating VEs using a wide-screen computer monitor was not found to significantly influence PAD ratings, perhaps due to insufficient participants/stimuli. In addition, considerable variation was seen by VE: e.g., for **Yumix Prada** (the VE inspired by Venice), there is quite a marked difference in favour of the HMD (which appears to have intensified the ratings already received when using a large monitor), whereas for **PickleSong** (haunted mansion in a forest), the HMD seems to have been somewhat less effective than the large computer screen.

Other observations were made based on data visualisations, but were unconfirmed by inferential models: in terms of Valence and Dominance scores, the Oculus Rift appears to have induced better discrimination between VEs, relative to when a large computer screen is used instead (i.e., VE average ratings were more spread out). In addition, presence/engagement levels were also found to be higher when the Oculus Rift was used - which visually had appeared to be the case especially for positive VEs, but without this being confirmed by models.

In terms of results from the statistical models created, the Oculus Rift was recognised by participants as being significantly more engaging than the large computer screen, as expected. However, the HMD also tended to be more associated with negative physical symptoms in this dataset, although this latter trend was not actually found to be significant in our models. Other findings from our models include that: no predictor was found to affect all three PAD dimensions simultaneously, and importantly, using an HMD over a large computer screen was never found to significantly alter any PAD ratings, be it for *Valence*, *Arousal* or *Dominance*.

Predictors that *were* successful in our models included the VE cluster (positive vs. negative), which significantly distinguished between *Valence* and *Arousal* scores, i.e., higher Valence was indeed associated with the positive cluster (as expected), similarly to lower Arousal. In contrast, the VE cluster was not a useful predictor for Dominance, suggesting the participants did not perceive the stimuli to vary enough in this regard. Indeed, the only two significant predictors for *Dominance* were: baseline Dominance (i.e., the higher the level of Dominance participants started the study with, the higher all VEs were rated on this measure), and Valence (i.e., the more positive the material was rated, the more dominant participants saw themselves in relation to it).

Returning to the prediction of *Valence* ratings, however, above and beyond VE type (or cluster), these were also predicted by: Presence/Engagement, Arousal and Dominance, so that the more engaging a VE was perceived, and the more dominant participants felt while exploring it, the more positively the VE was rated. In terms of other *Arousal* predictors (also above and beyond the VE cluster), baseline Arousal and Valence also contributed significantly to the model, so that: the more alert participants were at the onset of the study, the more arousing the VEs appeared to them; equally, the more negatively VEs were rated, the more arousing they were perceived to be (especially since the negative VEs included in this study were fear-based and high in Arousal). Importantly, no interactions were found between any of these predictors and the display output, so that these relationships apply equally to when a large monitor vs. an HMD was used to display the VEs.

In terms of predicting ITC factor scores, the only predictor to simultaneously affect both *Presence/Engagement* and *Physical symptoms* is Valence, in that: positive VEs were regarded as also more engaging (as a flipped logic from what was stated concerning Presence/Engagement predicting Valence scores), and negative VEs - as more likely to lead to negative physical symptoms. Separately, when predicting *Presence/Engagement*, other influencers included: the display output, with the large monitor being confirmed as significantly less engaging compared to the HMD; as well as Arousal scores, with more alerting VEs also being seen as more engaging.

As far as *Physical symptoms* are concerned, display output was not a significant predictor, i.e., while we know from visualising the data that VEs viewed with an HMD were associated with more physical discomfort, this appears to not have been enough to reach significance. Another risk factor for developing physical discomfort was baseline Arousal (i.e., participants starting the task in an alert state were more prone to experiencing discomfort), although two protective factors also emerged: baseline Valence and presence/engagement - in other words, beginning the task in a 'positive mood' and feeling present within the VEs made it less likely to feel physically unwell.

6.4.1 Limitations

Due to the inconsistent performance of the 6 VEs used as stimuli in this research, future studies should ideally employ a far wider variety of VEs. However, it is worth noting that this would also impose design restrictions (which were beyond the scope or resources available here). In order to avoid nausea and other harmful effects of prolonged HMD use, participants would need to undergo a multitude of study sessions - an option which is more costly, and more vulnerable to sample attrition over time. In addition, a series of technical limitations exist in the current study, which may have influenced the results presented above. With more advanced software and hardware becoming available over time, future research would be likely be less prone to the following concerns:

Kinetic feedback The kinetic feedback from the Rift's motion tracker can become decoupled from the direction of movement in the virtual world. This can lead to an unnatural feeling, with the sensation being similar to driving a motorcycle and moving one's head around: this allows to still see the surrounding scenes, but without being able to influence the direction of the vehicle this way. Particularly since one usually has a first person perspective when using the Rift with Second Life (i.e., only the environment is visible, not the avatar too), it is difficult to understand exactly how/when head orientation and direction of motion have become decoupled.

Control With the Rift on, environments with too many objects can be difficult to negotiate: it is often the case that without adequate control, one either bumps into objects, or gets so close to a wall that (due to a bug) one can see the environment beyond etc. All these issues can presumably decrease the sense of "realism" in Second Life.

Object interaction Object - avatar interaction can be challenging with the Rift, because the mouse location coordinates are decoupled between RiftMode and normal viewing. Whenever the mouse does become visible with the headset on, its apparent location seen with the HMD is not its real location, making it difficult to click and interact with objects. This means that each time the avatar needs to, e.g, open a door or get on a bus etc, the user would either use a trial-and-error process with clicking, or must drop out of RiftMode, use the mouse, and then return to RiftMode. Corrections for this are being considered[3].

Lag & simulator sickness There is some lag (less than 1s) between changing head position and actually triggering the corresponding change in camera angle in Second Life. This can be quite tiring, and simulator sickness symptoms (similar to motion sickness) may appear for this reason after varying lengths of time. This issue is related in part to hardware limitations, and also to the trade-off between rendering quality and rendering speed in Second Life. In this

[3] e.g., see David Rowe (2013), developer of CtrlAltStudio, http://ctrlaltstudio.com/blog/2013/09/01/alpha-2-of-the-ctrlaltstudio-viewer-with-oculus-rift-support: "Clicking on things at present is a bit fraught: the screen coordinates used by the viewer to work out where you click don't match up with where you click. This obviously needs to be fixed!"

study, given that the visual differences between low and high quality rendering are minimal in the Rift, we opted for a higher rendering speed, which *does* make a big difference: lag is less noticeable in this case, making the experience less uncomfortable. Future equipment (similar to the HTC Vive[4]) may be able to offer both simultaneously: excellent visual rendering, as well as great responsiveness and speed.

[4] https://www.vive.com/uk/

Part IV

Machine learning vs. human classification of stimuli

Chapter 7

Study 4: Using cluster analysis as a model for emotional categorisation in human participants

7.1 Introduction

THE link between concept categorisation and emotions can be traced back to the seminal work by Rosch, Mervis, Gray, Johnson, and Boyes-Braem (1976), who proposed that concept categorisations follow a hierarchical structure, with different levels of abstraction. The most general of these is the *superordinate category* level (e.g., "vehicle"), followed by the relatively more specific *basic categories* (e.g., "car" or "tank") which carry more information and are clearly distinguishable from one another, and finally, the *subordinate categories* (e.g., "limo", "scooter" etc.), which are the most specific.

Since then, research on the general categorisation process of concepts has been extrapolated to the categorisation of emotional information, giving rise to the prototype approach in emotions (Edelstein & Shaver, 2007; Shaver, Schwartz, Kirson, & O'Connor, 1987). According to this perspective, the superordinate level within the hierarchical taxonomy is represented by the concept "emotion", and is then further divided into basic categories, i.e., those used to distinguish across 'common' emotional states such as joy, anger, fear etc. The subordinate level is then populated by even more specific concepts such as rage, wrath or infatuation (Russell, 1991).

The association between categorisation processes and emotions has been taken even further, and incorporated explicitly even within the definition of a prototypical emotional episode, which "*refers to a complex process that unfolds over time, involves causally connected subevents (antecedent; appraisal; physiological, affective, and cognitive changes; behavioural response; self-categorisation), has one perceived cause, and is rare. Its structure involves categories (anger, fear, shame, jealousy, etc.) vertically organised as a fuzzy hierarchy and horizontally organised as part of a circumplex.*" (Russell & Barrett, 1999, p. 805). Other authors even consider conscious emotions in general to be a form of categorisation, where continuous components such as physiological signals, facial expressions and context are classified and labelled discretely using terms such as "anger" or "happiness" (Barrett, 2006). In fact, a failure to categorise emotional states in this way can be seen as a type of affective disorder (i.e., alexithymia, or the inability to identify and name emotional states, Bagby, Parker, & Taylor, 1994; Bagby, Taylor, & Parker, 1994; Parker et al., 2003; Taylor et al., 2003).

Other than their hierarchical organisation, an interesting property of emotion concepts and categories is that they are *fuzzy*, i.e., elements may not unambiguously belong to one category over another. Categories with fuzzy borders can explain, for instance, why there are difficulties / delays in deciding whether "pride" is an emotion or not (Russell, 1991). As a further illustration of these properties, Fehr and Russell (1991) studied the emotion of "love", as well as its hierarchical structure and fuzzy borders. They found that subtypes of love, such as affection, maternal love or self-love, could thus be ordered in terms of their *prototypicality* (i.e., depending on whether they are better or poorer examples of love, Edelstein & Shaver, 2007), with clearer examples of love being processed faster by participants when asked to verify their category membership (Brosch, Pourtois, & Sander, 2010).

Model-based cluster analysis (MBC), which we have used frequently throughout this work, reflects these properties of how (emotional) information is categorised (i.e., fuzzy category borders, and hierarchies of items varying in prototypicality). For instance, in MBC, cases are assigned probabilistically to clusters (which therefore also present fuzzy borders). Via MBC, cases can also be sorted in terms of their (un)certainty of cluster membership, hence forming a hierarchical structure where cases with a high certainty of membership are akin to category prototypes / best representatives (Fraley & Raftery, 2002, 2006). On the basis of these similarities between MBC and how humans categorise concepts/emotions, in this chapter we investigated whether it is possible to build an approximate model for how humans classify emotional information.

7.1.1 Aims

Because (emotion) concepts are organised hierarchically as *discrete* types (e.g., fear, subdivided into terror, anxiety, panic etc.), it has so far proved challenging to integrate variations in *continua* such as Valence and Arousal (Russell & Barrett, 1999) within a common model. However, MBC may propose a viable solution to this problem, by using these variations in PAD dimensions to estimate discrete groups/categories. To verify whether this is indeed the case, the current research examined how a human classification of affective VE stimuli may compare to a cluster analysis / machine learning classification of the same items.

More specifically, we set out to assess how similarly (if at all) a human classifier would group VEs into classes, compared to MBC, which suggested a $k = 2$ solution in Section 5.3.5 and Section 7.3.7. Any similarities (or differences) found could provide additional insight into how human participants form emotional categories around prototypical cases, with fuzzy borders. As far as the author is aware, the procedure employed here has not been used before, and may shed more light on how emotional categorisation processes unfold.

Finally, the research outlined in this chapter is a direct extension of our previous work in two ways: it refers back to the same clustering methods as before, and also samples virtual stimuli based on previously collected data (see Section 5.3.3 and Chapter 5).

7.2 Methods

7.2.1 Participants

Thirty volunteering participants were recruited for this study via the university Careers Service, and offered payment in exchange for their time commitment (£7.2/h). Their general characteristics are included below, in Table 7.1.

Table 7.1: Describing the sample in terms of gender and nationality.

Nationality	Female	Male
Africa	0	1
Eastern Europe	2	3
Far East	4	1
Middle East	0	1
North America	1	2
Western Europe	8	7

7.2.2 Stimuli

Stimulus selection for this study was based on data collected previously and described in Chapter 5 (p. 227). However, in the current study we implemented a different strategy for selecting stimuli, relative to previous chapters (see Chapter 6). Instead of simply averaging the (already-collected) raw PAD ratings by VE, and cluster analysing these average scores to form a sub-selection of stimuli, we improved this method by first *predicting* the raw PAD scores from a series of interfering covariates, and using only the resulting *residuals* within the cluster analysis, in order to remove unwanted variation from the data. Hence, Subsection 7.3.2 within the Results section discusses the process of how the virtual stimuli for this study were selected based on previous data: a factor analysing ITC-SOPI-SF items in order to create ITC factor scores, running a series of covariate models predicting PAD ratings from these ITC factor scores *as well as* other covariates, and then finally, extracting model residuals and clustering VEs based on them - with best representatives retained as stimuli for the current study.

Despite the fact that only VEs will be discussed as test cases here, other data were also collected: an identical experimental task was carried out on the 25 film stimuli mentioned previously in Chapter 5, with $N = 30$ participants. In addition, video data was collected for both stimulus groups, in order to record how the classification process of the stimuli occurred over time, e.g., if participants changed their minds concerning the number of existing categories, or what stimuli they included within these, and if there was any systematic pattern to these changes over time (a process we have named "category drift"). These data, however, exceed the scope of the current analysis, and shall not be discussed here.

7.2.3 Procedure

Participants were welcomed into the researcher's lab, and following a standard set of instructions (see Appedix F.1, p. 487), they began with a practice trial. The purpose of this was to demonstrate how both the Oculus Rift[1] and Second Life work, as well as how to input responses into the experiment builder program (i.e., OpenSesame). The practice phase was aimed at illustrating the structure of just one experimental trial. It relied on an example VE ('Rustica'[2]) which was unrelated to those included in the actual experiment. Participants were allowed to explore the VE for 1.5 minutes, with the possibility of extending this practice duration, if the procedure / commands were unclear. Extensions would not be granted in the actual experiment.

Next, participants filled in a series of questionnaire measures, and afterwards completed three tasks: they explored 25 randomised VEs, followed by rating these on both the PAD model and measures of presence/engagement, as well as classifying them into

[1] In the previous study comparing PAD ratings collected using a large monitor vs. the Oculus Rift, we found *ns.* differences between the Rift and the monitor. We proposed that the reason for this may be the large variation by VE, and the extremely small sample of VE stimuli used in the study. Hence, we still used the Rift in this study because, by testing more VEs here (i.e., 25) and measuring their PAD properties when displayed on an HMD, this new data could be compared to the monitor-only PAD ratings, which were collected for all 75 VEs used originally. This would serve as a verification of the previous interpretation of results. However, this is beyond the scope of the current study, and will form the objective of future investigations.

[2] http://maps.secondlife.com/secondlife/Rustica/183/119/52

groups however they saw fit, but based on their emotional content alone. The latter two tasks (i.e., providing ratings and classifying the VEs) were presented to participants in a counterbalanced order, depending on participant number parity. This measure was set in place to ensure that any error variance introduced into our data was minimised, and participants were at equal risk of: classifying the VEs to closely mirror how they rated them on the PAD dimensions, or alternatively, distorting spontaneous PAD ratings in favour of how they had classified the VEs beforehand.

On beginning the experimental task, participants were first asked to explore all 25 VEs, according to the following trial structure: within the experiment builder, one VE would be randomly selected from the list of 25. When ready, participants would click on a button and have this VE loaded into the separate Second Life viewer program. Also after the button had been clicked, participants would wait for 10s in front of a black screen while the online VE was loading. At the end of this brief period, window focus would shift from the experiment builder to the Second Life viewer automatically[3]. Participants would then know it was time to put on the Oculus Rift, and that a timer was set in motion for spending 1.5 minutes in the VE.

When the time allocated for exploring the VE had elapsed, an alarm would ring for 2s, and immediately afterwards, several actions ensued: the Second Life viewer teleported back to a 'baseline' VE, i.e., a virtual recreation of Edinburgh University (and in particular, the Informatics Forum[4]), and participants were given the option to have a break for 1.5 mins, which would be spent colouring in abstract geometric shapes, similarly to the previous study. However in this case, they were also allowed to skip the break in case they felt like no adverse physical symptoms had appeared following the use of the HMD. This decision was left to the participants at the end of each trial.

At the end of this VE presentation loop (which included the exploration of 25 VEs), participants would either proceed to the PAD rating task, or to the VE classification task, although both were completed by participants in counterbalanced order. In either case, VEs were rated or classified based on a series of *stills* - one generated for each VE. These captured the view available to participants on first entering each virtual world, and before moving to explore any of its specific scenery.

For the PAD rating task, each still was presented in random order for a duration of 3s in order to remind the participant of the VE to be rated. Afterwards, based on their experience in that world, participants were asked to rate "how [they] would feel if [they] experienced the previous situation in daily life", in terms of Valence, Arousal and Dominance[5]. Further questions were asked relating to presence and negative physical

[3] With the aid of: `win32gui`, `win32con`, `re`, `win32com.client`, and `time` Python libraries.

[4] http://maps.secondlife.com/secondlife/Edinburgh%20East/135/117/26

[5] These ratings were instrumental in assessing any overlap / discrepancy between the the spontaneous human classifications, and a parallel clustering solution of the same stimuli, created based on the PAD ratings alone.

symptoms, using items from the ITC-SOPI-SF.

When starting the classification task, a VE reminder menu would appear in OpenSesame. This was a 5×5 grid of VE thumbnails (or smaller versions of the stills mentioned above) which is illustrated in Figure 7.1 (p. 297). In case participants could no longer recognise a world they visited based on the thumbnail alone, this reminder menu was available to offer a refresher: clicking any reminder thumbnail in the OpenSesame grid would teleport the participant back into the respective world with HMD mode automatically engaged - but for only $1/3$ of the original exploration time, i.e., 30s. Participants were afterwards returned to the usual view, where they could alternate once more between the classification task (occurring in a browser), and the reminder menu.

Participants were instructed to intuitively and spontaneously create groups of stimuli - but only based on the emotional states they experienced while in those worlds. Two particularities marked this task: namely, the classification of VEs was completely undirected, i.e., the number of groups, the number of elements within groups, and how the groups were labelled was left up to each participant to decide. However, one *a priori* category was offered to participants, i.e., a group of 'Unclassifiable' VEs, which they could choose to populate or not, depending on whether they felt that a few stimuli (dis)agreed with their classification structure to any significant extent. Finally, participants were also asked to rank the elements placed within each category according to how 'representative' they were for the group - with the top item being the most representative, and the bottom item, the least representative. The interface for this classification task is shown in Figure 7.2 (p. 298), alongside more detailed explanations.

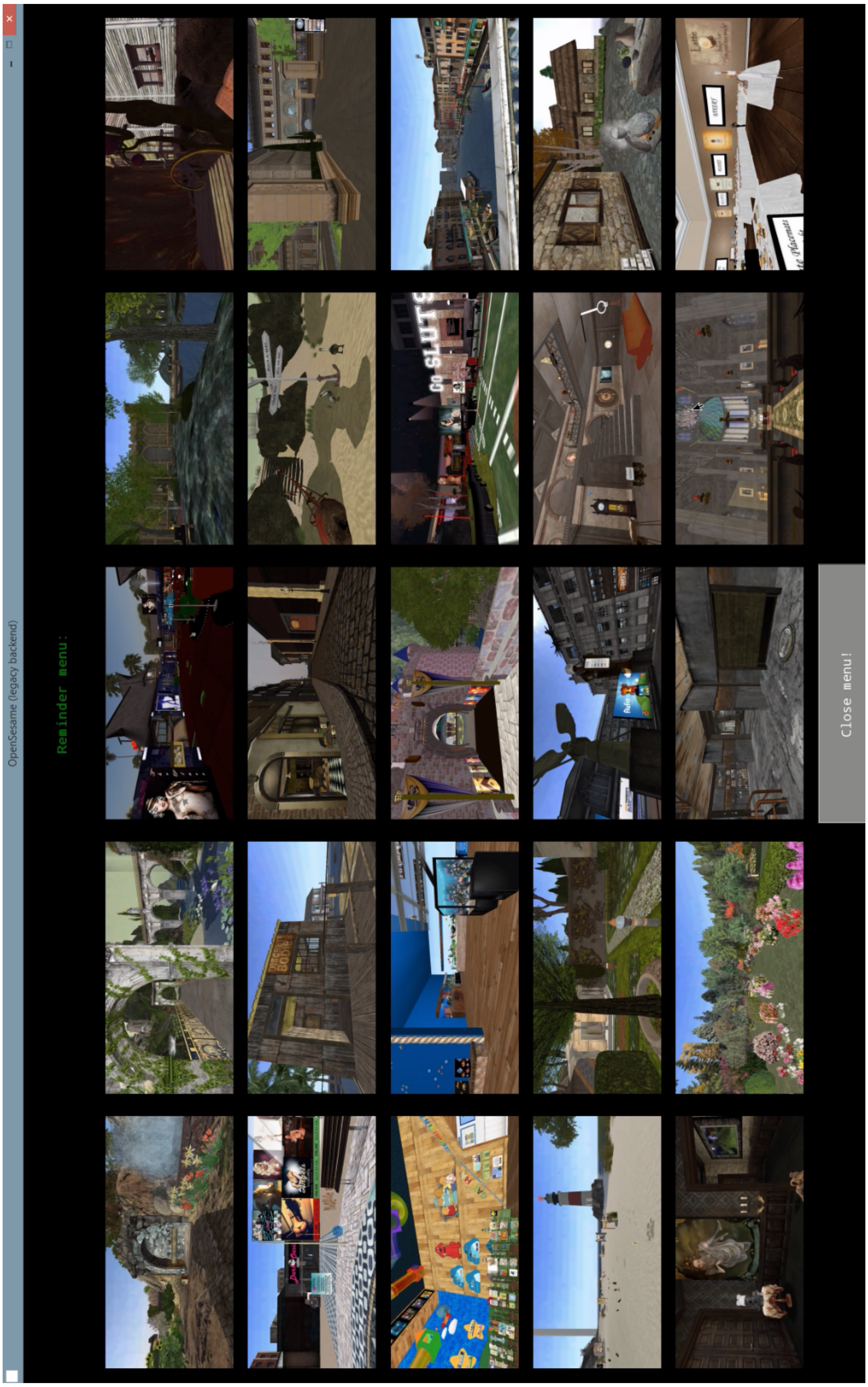


Figure 7.1: Reminder menu in OpenSesame. The configuration of the 25 thumbnails within the grid was random for each participant. On clicking any thumbnail, the participant would be teleported to the corresponding VE for 30s, as an on-demand refresher for their emotional experience.

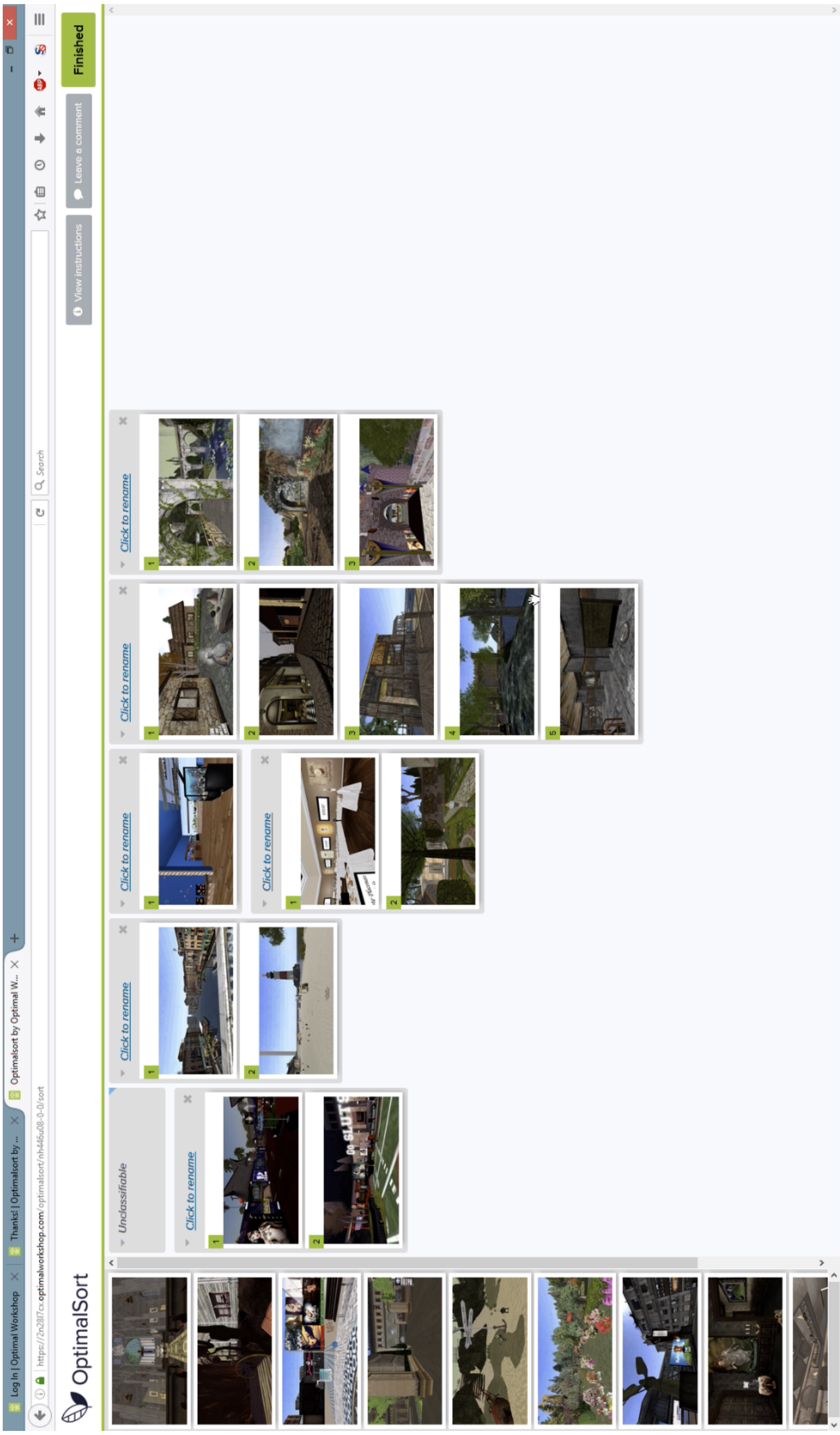


Figure 7.2: OptimalSort interface, where the cards in the leftmost column represent the VEs which all participants will have visited. During this sorting task, participants were asked to gradually drag the VE thumbnails from the leftmost column onto the wide empty space to the right. Categories are formed by 'chaining' the thumbnails and labelling their group. Participants were also asked to make sure that the order in which they sort the thumbnails is indicative of how representative they believe each VE to be within their category, with the top sorted VE being the best representative for the category. Finally, in this *hybrid* sorting task, all participants were offered a pre-made and optional **'Unclassifiable' category**, where they could store any VEs which they believed did not fit with the rest of their classification structure.

7.2.4 Apparatus

A suite of programs was jointly used to run this study. The three main programs used were: OptimalSort[6] - an online platform for creating classification (or 'card sorting') tasks for participants (among others such as 'tree testing' for validating website menu structure, qualitative research etc.). As mentioned before, the card sorting interface of OptimalSort is shown in Figure 7.2 (p. 298), alongside a description for how the entire system can be used.

Secondly, we again used CtrlAltStudio to display the Second Life virtual environments used as stimuli (with output directed to the Oculus Rift DK1). The use of CtrlAltStudio was coordinated with a third program - the OpenSesame experiment builder. The latter was required to randomise the presentation of VEs, as well as capture participant ratings (since PAD ratings and questionnaire answers were also captured alongside the VE classification within OptimalSort).

Finally, we also employed the Atomi Active Presenter desktop recorder, in order to have a record of participants' behaviour within the VEs, and of how many solution iterations each participant went through before settling on their final VE classification. For the former purpose, this measure was set in place to ensure participants did not stray from the task in the VEs (e.g., refused to respect the time limit imposed, or interacted with other avatars even if their sole task was to explore the environments etc.). A few of these video materials were randomly inspected at the end of the data collection period, and were found to conform to the instructions provided by the researcher. The second purpose for using a desktop recorded was so that, at a later date, it would be possible to return to these video logs and measure 'category drift' over time.

7.2.5 Instruments and measures

The same instruments were used in this study, as for all the previous research in this thesis. The only novel measures included (owing to the introduction of OptimalSort) were information on: the classification of virtual stimuli into groups, and on the hierarchy of elements within these groups (according to how 'prototypical' participants judged them to be for the category in which they were placed).

[6] https://www.optimalworkshop.com/optimalsort

7.3 Results

7.3.1 Sample descriptives

The following results in Table 7.2 were computed using `R`, and the R packages listed in Appendix section F.4 (p. 525).

Table 7.2: Sample description in terms of age, media usage, the PANAS schedule, PHQ, TAS-20, and baseline Valence, Arousal and Dominance, for $N = 30$. Interestingly, all participants rated themselves the same on the `CompFreq` measure, hence SD = 0.

No	Measure	Mean	Trim	Median	SD	Min	Max	Range	Skew	Kurt	SE
1	Age	21.93	21.58	21.00	3.36	18.00	31.00	13.00	0.88	-0.07	0.61
2	CompComf	3.80	3.92	4.00	0.48	2.00	4.00	2.00	-2.25	4.38	0.09
3	CompFreq	4.00	4.00	4.00	0.00	4.00	4.00	0.00	NaN	NaN	0.00
4	FilmFreq	2.77	2.75	3.00	0.57	2.00	4.00	2.00	-0.01	-0.51	0.10
5	GameFreq	2.00	2.04	2.00	1.14	0.00	4.00	4.00	-0.27	-0.86	0.21
6	MMORPG Freq	0.63	0.46	0.00	0.93	0.00	4.00	4.00	1.75	3.38	0.17
7	PhotoFreq	2.80	2.88	3.00	1.03	1.00	4.00	3.00	-0.53	-0.91	0.19
8	SLFreq	0.07	0.00	0.00	0.25	0.00	1.00	1.00	3.30	9.21	0.05
9	VWFreq	0.57	0.54	1.00	0.57	0.00	2.00	2.00	0.29	-1.02	0.10
10	PANAS_Neg	5.10	4.92	4.50	2.56	1.00	11.00	10.00	0.55	-0.29	0.47
11	PANAS_Pos	13.03	13.33	14.00	2.74	4.00	17.00	13.00	-1.24	1.75	0.50
12	PHQ_total	4.43	4.17	3.50	3.14	0.00	12.00	12.00	0.61	-0.60	0.57
13	TAS20_DIF	6.10	5.17	5.00	6.30	0.00	23.00	23.00	1.08	0.22	1.15
14	TAS20_DDF	6.93	6.62	6.00	3.69	2.00	15.00	13.00	0.61	-1.02	0.67
15	Base V	5.67	5.71	5.50	1.40	3.00	8.00	5.00	-0.15	-0.65	0.26
16	Base A	2.23	2.08	2.00	1.77	0.00	7.00	7.00	0.62	-0.26	0.32
17	Base D	4.37	4.33	4.00	1.19	2.00	7.00	5.00	0.37	-0.66	0.22

Note. **Base A** = Baseline Arousal; **Base D** = Baseline Dominance; **Base V** = Baseline Valence; **CompComf** = level of comfort when using a computer; **CompFreq** = frequency of using a computer; **FilmFreq** = frequency of watching films; **GameFreq** = frequency of gaming; **MMORPGFreq** = Massively Multiplayer Online Role-Playing Games gaming frequency; **PANAS_Neg** = PANAS Negative Affect scale; **PANAS_Pos** = PANAS Positive Affect scale; **PhotoFreq** = frequency of taking photos; **PHQ_total** = PHQ-8 total score; **SLFreq** = frequency of using Second Life; **TAS20_DIF** = TAS-20 Difficulty Describing Feelings subscale; **TAS20_DIF** = TAS-20 Difficulty Identifying Feeling subscale; **VWFreq** = frequency of using virtual worlds.

7.3.2 Selecting a subset of 25 virtual stimuli from the previous VE group ($N = 51$)

7.3.2.1 Reverification of ITC-SOPI-SF dimensionality

In order to select 25 stimuli, we first re-ran the factor analysis described in Section 4.3.5, as a means of producing ITC factor scores - which would later serve as covariates. The same factor structure re-emerged, with two components / factors underlying the data (suggested by a parallel analysis, Dinno, 2009), rather than the four expected by the original authors (Lessiter et al., 2001). Results of Horn's Parallel Analysis for factor retention (as computed using the R package and function `paran::paran`, with 5000 iterations) suggest that 2 factors should be retained, the first with an adjusted and unadjusted eigenvalue of 1.699 and 1.753, respectively, and the second with 0.074 and 0.097, respectively.

These factors were subsequently extracted using both Varimax (orthogonal) rotation and Oblimin (oblique) rotations, which did not affect results (i.e., loadings) to any important degree due to a low inter-factor correlation, i.e., $r = 0.075$. See Table 7.3 below for further results:

Table 7.3: Loadings of the ITC-SOPI-SF items on the two factors extracted, with Oblimin and Varimax rotations.

Rotation	Item	Factor 1 loadings[a]	Factor 2 loadings[b]	Communality[c]	Uniqueness[d]	Complexity[e]
Oblimin	ITC1	0.87	-0.10	0.75	0.25	1.0
	ITC2	0.85	0.00	0.73	0.27	1.0
	ITC3	0.63	0.27	0.50	0.50	1.4
	ITC4	-0.15	0.39	0.17	0.83	1.3
Varimax	ITC1	0.82	-0.27	0.75	0.25	1.2
	ITC2	0.84	-0.18	0.73	0.27	1.1
	ITC3	0.69	0.13	0.50	0.50	1.1
	ITC4	-0.04	0.41	0.17	0.83	1.0

[a] Proportion of variance explained by Factor 1, with Oblimin rotation: 0.48, and with Varimax rotation: 0.47.

[b] Proportion of variance explained by Factor 2, with Oblimin rotation: 0.06, and with Varimax rotation: 0.07.

[c] Communality refers to the proportion of variation in that item explained by the two factors.

[d] Uniqueness is 1 - communality.

[e] Complexity is the extent to which items cross-load on multiple factors.

Based on these results, we created a summed score for Factor 1, i.e., Presence/Engagement (across items ITC1, ITC2 and ITC3), and retained ITC4 as a representation of Factor 2, i.e., Physical Symptoms. After creating ITC factor scores and using these as covariates alongside others (to be mentioned shortly), error variance due to these covariates should be excluded from the original data - and therefore, only the 'useful', unexplained variance would subsequently used to classify and select the VEs for this study. As previously, once each case in the original dataset was associated with a residual value from these covariate models, the residuals were averaged by VE, and it is these averaged residuals that were fed to a new cluster analysis.

7.3.2.2 Selecting VE stimuli based on covariate models predicting PAD dimensions

In order to obtain residual values expressing only ‘relevant’ PAD variance which can then be clustered, we created three models: one predicting VE Valence scores, another for Arousal, and finally a third for Dominance ratings. Each of these included a constant set of predictors (covariates), which appear below in Listing 7.1:

Listing 7.1: `R` code snippet: General form for the `lmer()` models used to Valence, Arousal and Dominance scores, and extract residuals for subsequent clustering. Random intercepts were included by participant and VE (i.e., ‘`New.SLURL`’), and all continuous predictors were scaled, as usual.

```r
# General structure for lmer() models:
s <- function(x){ scale( x, center = TRUE, scale = TRUE ) }

full_Valence_model <- lmer( rateV ~
                              s(similarity) +
                              s(rateBaseV) + s(rateBaseA) + s(rateBaseD) +
                              as.factor(session) +
                              s(Age) + Gender + Nationality +
                              s(PhotoFreq) + s(FilmFreq) + s(CompComf) +
                              # s(CompFreq) + # This breaks the computation, and was excluded because every person gave a rating of 4. SD = 0.
                              s(GameFreq) + s(MMORPGFreq) +
                              s(VWFreq) + s(SLFreq) +
                              as.factor(stimulusSoundFilmVRMusic) +
                              s(PANAS_Neg) + s(PANAS_Pos) + s(PHQ_total) + s(TAS20_DDF) + s(TAS20_DIF) +
                              s(ITC_Presence_Factor) + s(ITC_PhysicalSymptoms_Factor) +
                              (1 | subject_nr) +
                              (1 | New.SLURL),
                            data = DataForResiduals,
                            verbose = TRUE, REML = FALSE,
                            control = lmerControl( optimizer = "bobyqa",
                                                   optCtrl = list( maxfun = 50000 ) ) )
```

Table 7.4: Statistical models: Single predictor found for Valence, Arousal and Dominance scores of all those tested (and specified above in Listing 7.1).

	Valence	Arousal	Dominance
(Intercept)	4.32***	3.89***	4.16***
	(0.17)	(0.18)	(0.14)
ITC Presence/Engagement Factor	0.33***	0.23***	0.22***
	(0.04)	(0.05)	(0.04)
AIC	10879.77	11392.67	10618.33
BIC	10909.64	11422.53	10648.19
Log Likelihood	-5434.89	-5691.33	-5304.17
Num. obs.	2900	2900	2900
Num. groups: Participant	59	59	59
Num. groups: SLURL	51	51	51
Var: Participant (Intercept)	0.22	0.87	0.54
Var: SLURL (Intercept)	1.16	0.77	0.48
Var: Residual	2.26	2.66	2.05

$^{***}p < 0.001$, $^{**}p < 0.01$, $^{*}p < 0.05$

SLURL = Second Life URL, or VE identifier

Within these models, all continuous predictors were scaled, and model estimates were chosen to optimise log-likelihood values, rather than the restricted (residual) maximum likelihood (REML)[7], in order to be able to compare models based on AIC fit criteria. This comparison was achieved using backwards, stepwise selection, which revealed, interestingly, that the only covariate which was useful to control across all three outcomes was the ITC Presence/Engagement Factor scores (with results shown in Table 7.4). It is worth noting that VE cluster membership was *not* used as a categorical predictor in the models above (hence, neither was it included as a random effect), because this would have removed 'valuable' variance as well - whereas the purpose of these models was only to discard any 'error' variance due to the covariates[8]. Once these residuals were computed for each PAD dimension / model, they were aggregated by VE regions (i.e., SLURL) and then clustered using model-based clustering in `R`.

Based on the residuals for each PAD dimension and a model-based cluster analysis, the 51 VEs were grouped into an optimal structure of $k = 2$, where the components were of ellipsoidal shape, equal volume and orientation (i.e., an EVE model), with BIC = 671.156. Two other competing models were also for $k = 2$: an EVV model (BIC = 667.986), and a VVE model (BIC = 667.747). For details on the classification and

[7] Hence, REML was switched to '`FALSE`' in `R` models and syntax.

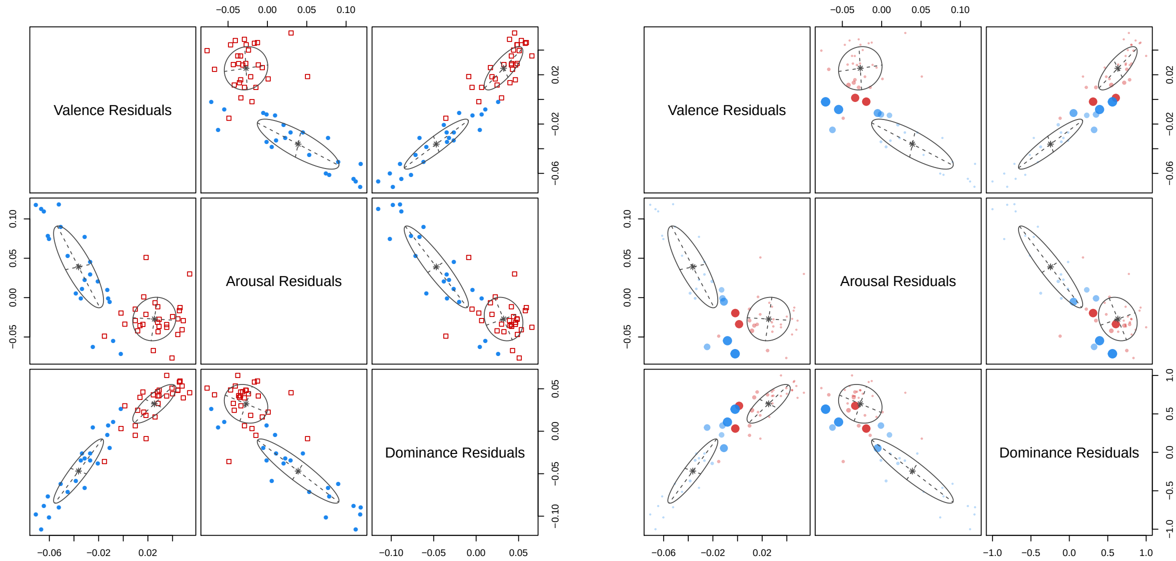
[8] As a verification for this, indeed inserting cluster membership as a predictor / fixed effect (alongside its associated random slopes) led to residuals that could only be clustered into one cluster, i.e., no clusters could be identified in the data any longer

undertainty of individual VEs, see Table 7.5 and Figure 7.3 (p. 305):

Table 7.5: MBC classification and uncertainties for the 51 VEs also used in a previous study and described in Table 5.11. Here, however, the MBC solution has been achieved by using Valence, Arousal and Dominance residuals, rather than raw scores.

No	SLURL	Uncert.	Cluster	No (cont.)	SLURL	Uncert.	Cluster
1	Agriopis	0.159	1	27	Kismet Northwinds	0.000	2
2	Alexandre	0.000	2	28	Left Hand Column	0.050	2
3	Alice	0.034	2	29	Mediterraneo OC	0.023	2
4	alirium	0.001	2	30	Misthaven	0.016	2
5	Angel Manor	0.000	2	31	Mlastina de Anticii	0.000	1
6	Aquitaine Coeur Nord	0.001	2	32	Mountains of Creta	0.000	2
7	Baraka Point	0.000	1	33	New Toulouse Bayou	0.000	2
8	Bracket	0.001	2	34	Oceanea	0.000	2
9	Brandy Wine Island	0.000	2	35	Picklemoon	0.000	1
10	BROTHEL	0.001	1	36	PickleSong	0.000	1
11	Burning Hart	0.000	2	37	Pino	0.000	2
12	Calypso Reef	0.002	2	38	Port Babbage	0.003	2
13	Canis Beach	0.000	2	39	PREFABRICA	0.006	1
14	Cavettaz	0.015	2	40	Quietly Tuesday	0.000	2
15	Depoz Specialties	0.000	2	41	Second Health London	0.422	1
16	Dreyfus	0.314	2	42	Sexy Sands1	0.098	1
17	escort oasis	0.001	1	43	Trianwe	0.167	1
18	FireStorm	0.002	1	44	Triglav	0.011	2
19	Furniture	0.450	1	45	TT Enterprises	0.001	2
20	Furor	0.000	1	46	Tulagi	0.014	1
21	GUER-REIROS	0.001	1	47	Turia	0.011	2
22	Harshap	0.000	1	48	Twilight Hollow	0.000	1
23	Hell	0.000	1	49	Weedon Island	0.000	1
24	Isle of Tharen	0.004	1	50	Xalfor	0.283	1
25	Kalepa	0.000	2	51	Yumix Prada	0.000	2
26	Kindred Spirit	0.335	2				

As a reminder, in our previous work (see Section 5.3.3, p. 243), we devised a method



(a) Clustering VE residuals.

(b) VE uncertainties after clustering residuals.

Figure 7.3: VE clusters using model-based clustering with PAD residuals. Two clusters emerge and mimic the structure found previously in Figure 5.13 (p. 255). These clusters largely represent general positive and negative affect.

for excluding VEs which changed too drastically over time. In this fashion, only the 51 VEs above remained viable for use in analyses, of the total number of 75. However, at the time of writing, several of these 51 VEs had since either been modified, sold or deleted, and hence also needed to be excluded from further analysis. Those that remained were filtered further based on their uncertainties, so that 7 VEs presented the highest certainty of belonging to general positive cluster, and another 7 - to the general negative cluster. From the remaining VEs, a further 11 were selected to sample the space of uncertainty in between these clusters, in this manner: 11 equally spaced cut-off points were created on a distribution of probabilities from 0 to 1, i.e.: 0.10, 0.19, 0.28, 0.37, 0.46, 0.55, 0.64, 0.73, 0.82, 0.91, 1.00. Based on the distribution of VE uncertainties (also varying between 0 and 1), we then found the nearest values corresponding to these 11 cut-off points / probabilities. Then, we simply included the corresponding VE into our selection of stimuli. This resulted in a list of 11 'uncertain' VEs, which, along with the previous 7 positive, and 7 negative VEs, presented the following characteristics specified in Table 7.6:

Table 7.6: Listing of 25 VEs used in the current classification study.

No	SLURL	Group Label	Uncertainty	Cluster
1	Picklemoon	Negative	0.000000	1
2	PickleSong	Negative	0.000000	1
3	Harshap	Negative	0.000003	1
4	escort oasis	Negative	0.000766	1
5	BROTHEL	Negative	0.000923	1
6	GUERREIROS	Negative	0.001384	1
7	PREFABRICA	Negative	0.005686	1
8	Kalepa	Positive	0.000000	2
9	Yumix Prada	Positive	0.000000	2
10	Quietly Tuesday	Positive	0.000000	2
11	Brandy Wine Island	Positive	0.000000	2
12	Angel Manor	Positive	0.000001	2
13	Mountains of Creta	Positive	0.000002	2
14	Kismet Northwinds	Positive	0.000005	2
15	Depoz Specialties	Uncertain	0.000085	2
16	Canis Beach	Uncertain	0.000262	2
17	Bracket	Uncertain	0.000558	2
18	Port Babbage	Uncertain	0.003295	2
19	Triglav	Uncertain	0.011209	2
20	Cavettaz	Uncertain	0.015312	2
21	Misthaven	Uncertain	0.015944	2
22	Mediterraneo OC	Uncertain	0.023141	2
23	Alice	Uncertain	0.034087	2
24	Agriopis	Uncertain	0.159294	1
25	Trianwe	Uncertain	0.167110	1

7.3.3 Scanning VEs along the duration of the study

Over the course of this study, VE similarities were computed in a manner already described in Section 5.3.3, p. 243. Inventory scans were run bi-weekly, adding up to a total of 10 scans, dated between 2016-10-10 and 2016-10-31. In summary, when two inventory scans were being compared for the same VE, we measured the level of similarity between them as being: the number of identical items simultaneously occurring in both scan dates, divided by the total number of pooled items across both scans, and with the result multiplied by 2 (for cases when the scans being compared were actually identical). This led to normalised results ranging from 0 to 1, where a value of 0 represents zero similarity (i.e., that the VE changed completely between dates), and 1 represents perfect similarity (i.e., no changes between scans).

As discussed previously, selecting two scans for comparison involved creating pairings between every possible 'reference' date, and every possible scan date. As such, it was necessary to first determine which reference scan maximised similarities between scans. In order to do this, in a data matrix that is available for inspection in Appendix Section F.2, we computed row means for each combination of VE and reference date, thus leading to 25 averaged similarity values per reference date (one for each VE). It is these values that are shown in the boxplots from Figure 7.4. With the exception of some outliers, the figure shows the reference scans to be performing quite similarly. By further collapsing the 25 data points into a single, averaged similarity value for each reference date, Figure 7.5 shows the optimal reference date to be 2016-10-21, by a very narrow margin (similarity = 0.9202).

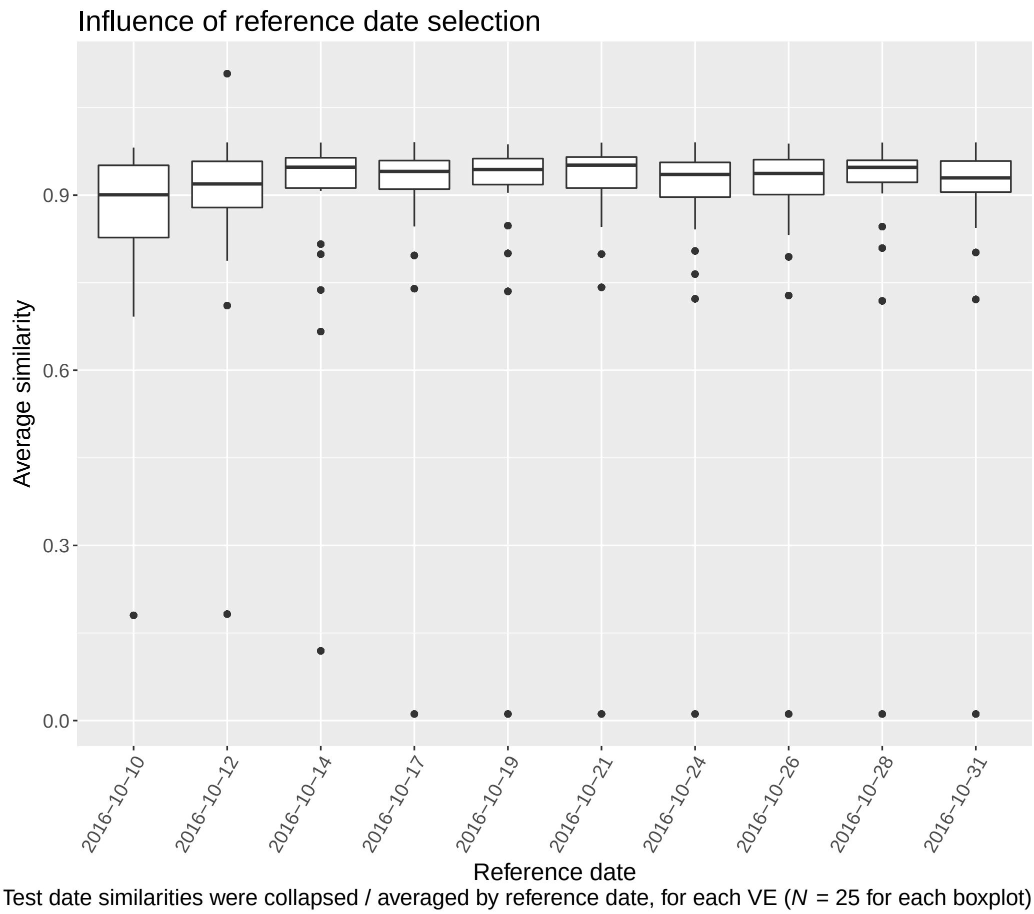


Figure 7.4: Boxplots showing average similarities per VE, across test dates. Each boxplot shows the distribution of average similarities achieved by each reference scan, for all 25 VEs. With the exception of some outliers, in this case the reference scans seem to perform relatively similarly.

Having selected an appropriate reference date, we then used it to inspect more closely the patterns of change which the VEs went through. They are displayed in the heatmap from Figure 7.6 (p. 309) below, and show that, with the exception of the Brandy Wine Island VE (which was excluded from subsequent analyses because it became unavailable

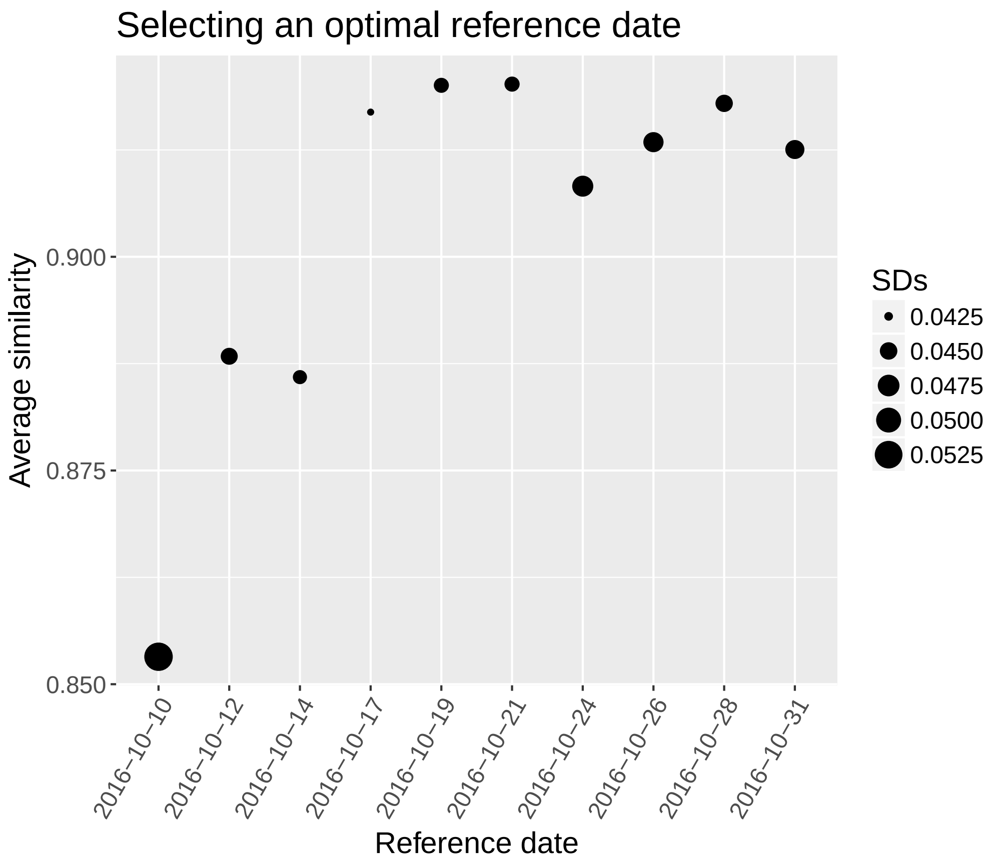


Figure 7.5: Scatterplot showing average similarities per reference date, across test dates *and* VEs. Essentially, each point is a summary for the N = 25 data from each boxplot in Figure 7.4. The size of points is proportional to SDs derived from the same 25 individual VE similarity values, which went into the creations of the means displayed.

shortly after the start of the study), all the other VEs were relatively consistent over time. Hence, all 24 remaining VEs were considered to vary insignificantly over the duration of the study - so much so, that this was not regarded as a threat to inference.

7.3.4 VE user classification tendencies

On average, participants created 7.17 categories (SD = 2.31). More information about participant classification tendencies is presented visually in Figure 7.7 (p. 310), which also suggests that the number of VE groups created by participants tends to be higher for VEs with higher uncertainties. Hence, if participants are unsure how to sort such VEs, statistical uncertainties may indeed translate into how humans perceive emotional information, and induce an inflation in VE categories. It is equally the case that the number of 'uncertain' VEs used in the study also happened to be higher ($N = 11$) than the number of 'clearly positive' ($N = 6$) or 'clearly negative' ($N = 7$) VEs - and this too could have resulted in the uncertain VEs finding themselves scattered into more

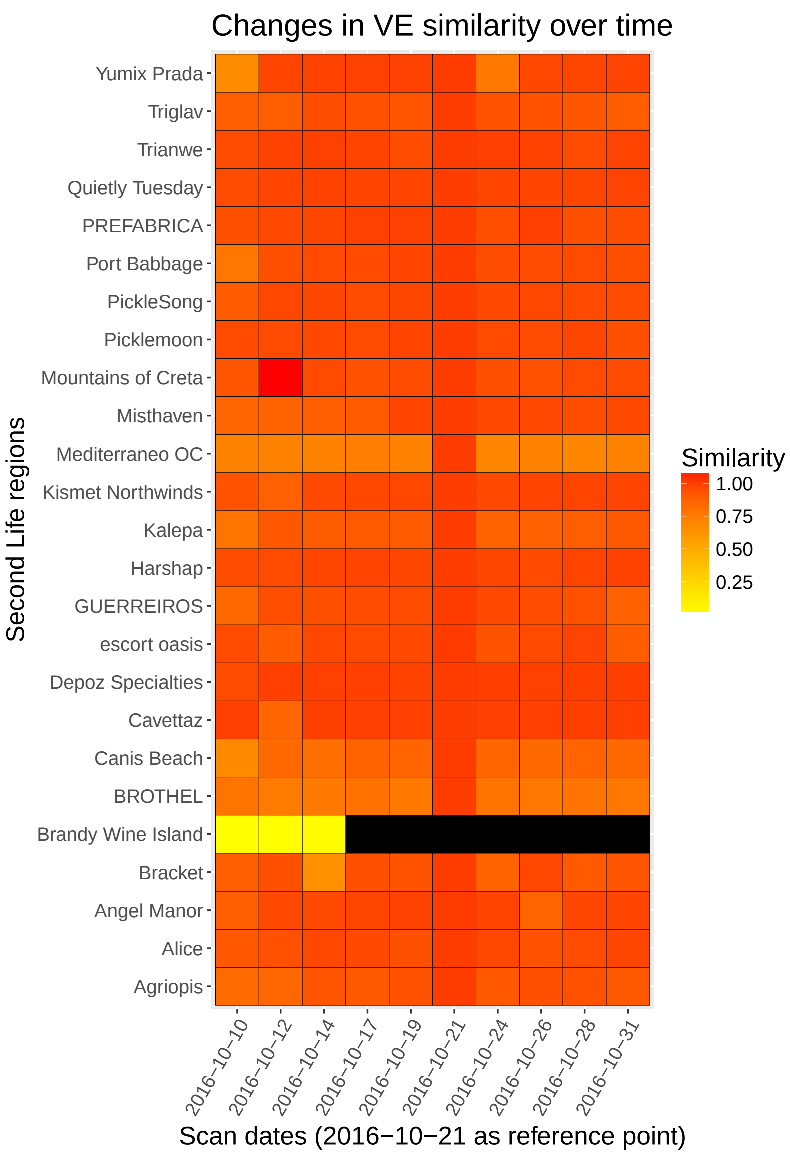
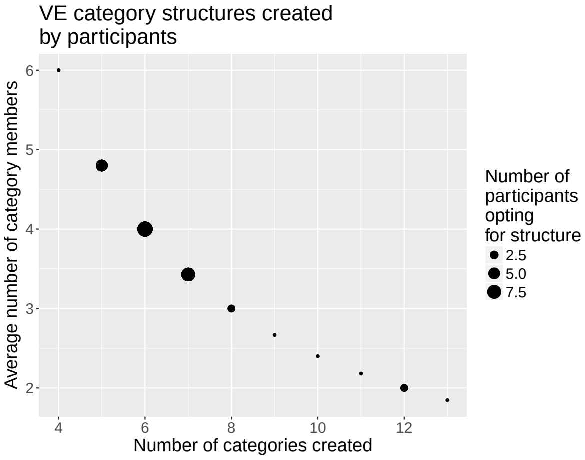


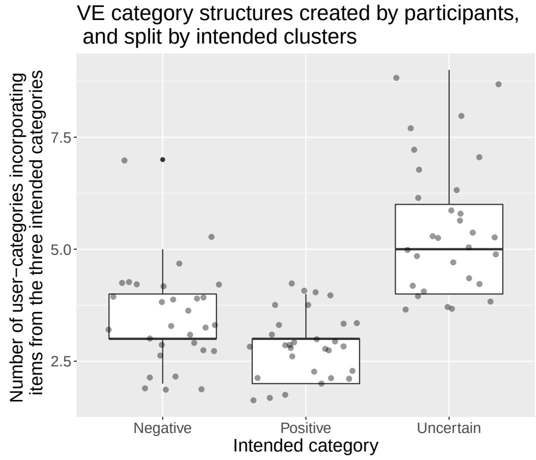
Figure 7.6: Heatmap for the 25 VEs and their associated similarities, as assessed using "2016-10-21" as the reference date. Brandy Wine Island shows the poorest consistency, and even vanishes from Second Life before the fourth scan even occurs (black cells represent missing data). Hence this VE was excluded before further analyses, leading to a $N = 24$ VEs.

user-categories, than the positive or negative ones.

To distinguish between these two possibilities, i.e., if it was indeed the higher level of uncertainty for these VEs, or merely their larger number which led to their inclusion among more (and finer-grained) categories, relative to the positive or negative VEs, we created a linear mixed-effects model which predicted the number of VEs within each user-made category (i.e., a measure of "category size" across participants, which would be a larger number for few, coarse-grained categories, and a smaller number, for many fine-grained categories), from: VE uncertainties, the type of VE (uncertain, positive or



(a) VE category structure



(b) VE category structure by intended cluster. The boxplots show how many user categories each of the intended categories was spread across.

Figure 7.7: VE category structure. Most participants created 5, 6 or 7 categories overall. Also, the number of categories which users spread the 'uncertain' items across is indeed higher than for the "clearly positive" or "clearly negative" items (as identified using MBC). For these latter two types, structures with fewer, broad categories dominated. Hence, statistical measures of uncertainty may indeed translate into how humans perceive emotional information.

negative), and the interaction of the two.

Results are presented in Table 7.7, and include a main effect of VE uncertainty on category size: the higher the level of MBC uncertainty, the larger the category created by participants. Interestingly, this pattern is reversed for VEs in the 'uncertain group' only (i.e., those sampled from the space in between the two clusters: positive and negative). In their case, rather, higher uncertainty is associated with the creation of smaller, finer-grained user categories. Hence, it is unlikely that this pattern of results is simply due to there being more uncertain VEs available for participants to sort, compared to the other types.

Table 7.7: VE uncertainty mixed model, predicting category size from uncertainty measures and VE type. Uncertainty values were scaled, and random intercepts were added by VE, as well as participant - with the latter allowed to vary by VE type.

	VE uncertainty model
(Intercept)	6.53^{***}
	(1.36)
Scaled uncertainty values	7.02^{*}
	(3.50)
VE type: positive	902.73

	(1623.51)
VE type: uncertain	-2.30
	(1.36)
Scaled uncertainty × VE type: positive	2196.26
	(3956.29)
Scaled uncertainty × VE type: uncertain	-7.01^{*}
	(3.50)
AIC	2655.95
BIC	2720.06
Log Likelihood	-1313.97
Num. obs.	720
Num. groups: Participant	30
Num. groups: VE	24
Var: Participant (Intercept)	1.18
Var: Participant × VE type: positive	1.72
Var: Participant × VE type: uncertain	0.68
Cov: Participant (Intercept) × VE type: positive	-0.44
Cov: Participant (Intercept) × VE type: uncertain	-0.16
Cov: Participant × VE type: positive × VE type: uncertain	1.01
Var: VE (Intercept)	0.09
Var: Residual	1.89

$^{***}p < 0.001$, $^{**}p < 0.01$, $^{*}p < 0.05$

7.3.5 Co-occurrence analysis

Because participants were allowed to group VEs in whichever way they saw fit (i.e., using any number of categories, with any verbal name/label, and with any number of members), what was of particular interest in this study was the *pattern of co-occurrences* in the data - in other words, which VEs tended to be placed together in the same category - regardless what the category was named, how many (other) items were in it, or how many categories existed overall.

To investigate this, we borrowed a statical model typically used in ecology research for studying species co-occurrences within sites, and which is implemented in the `cooccur` `R` package (Griffith, Veech, & Marsh, 2016). In brief, according to the authors, co-occurrence is often measured as the number of ecological sites (i.e., participant-made categories, in this case) where two species (i.e., VEs, here) have co-occurred. By contrasting the co-occurrences of species within sites (VEs within participant categories) which were observed vs. expected by chance, it is possible to determine the significance level for *positive* co-occurrences (when species appear together in sites more frequently than expected), and *negative* co-occurrences (when species appear together less than would be expected by chance). This model will also help to identify *random* co-occurrences, or cases where observed co-occurrences did not differ from those expected. Furthermore, the `cooccur` package also interprets co-occurrences as the probability of selecting a site that presents species A, given that it already presents species B.

Because the data associated with this study have a nested structure unlike the ecology analogy above (i.e., VEs within categories within participants, more similar to species within sites within countries), we created binary presence / absence matrices for each participant, coding in 0s and 1s whether or not a given VE had been placed within a given stimulus category. At the end of this process, we simply stacked the individual participant presence / absence matrices, and used this binary dataset to investigate co-occurrence patterns.

By running a co-occurrence analysis, we discovered that, for a total of 24 species (i.e., VEs), and 215 sites (i.e., user-defined VE categories, expanded across all participants), the VEs formed a total of 276 analysable pairs, of which: 43 were significant positive associations, 38 were significant negative associations (i.e., overall, 29.3% non-random co-occurrences, at $\alpha \leq 0.05$), and 195 were random pairs (see Figure 7.8, p. 315). These associations are also displayed in Figure 7.9 (p. 316), which further highlights the existence of a positive and negative "node" or pole of paired VEs, all of which tend to occur frequently together within the same participant-made categories. The 43 significant positive associations (which were considered of higher interest than the negative co-occurrences) are listed in Table 7.8, which is accompanied by further details.

Table 7.8: VE significant positive co-occurrences, which very often occur within the same MBC cluster / VE type. VE cluster membership is based on previously collected PAD data - which directed stimulus selection before the current study was deployed.

No	VE 1	VE 1 Cluster Label	VE 1 Cluster	VE 2	VE 2 Cluster Label	VE 2 Cluster	Obs. co-occurrences	p	Effect size
1	Brothel	Negative	1	Escort Oasis	Negative	1	27	0.000	0.106
2	Brothel	Negative	1	Guerreiros	Negative	1	27	0.000	0.106
3	Escort Oasis	Negative	1	Guerreiros	Negative	1	26	0.000	0.101
4	Pickle-moon	Negative	1	Picklesong	Negative	1	17	0.000	0.060
5	Harshap	Negative	1	Pickle-moon	Negative	1	10	0.003	0.027
6	Harshap	Negative	1	Picklesong	Negative	1	10	0.003	0.027
7	Harshap	Negative	1	Prefabrica	Negative	1	10	0.003	0.027
8	Kalepa	Positive	2	Kismet North-winds	Positive	2	14	0.000	0.046
9	Kalepa	Positive	2	Mountains Of Creta	Positive	2	17	0.000	0.060
10	Kismet North-winds	Positive	2	Mountains Of Creta	Positive	2	14	0.000	0.046

11	Kismet North-winds	Positive	2	Quietly Tuesday	Positive	2	13	0.000	0.041
12	Angel Manor	Positive	2	Quietly Tuesday	Positive	2	12	0.000	0.036
13	Kalepa	Positive	2	Quietly Tuesday	Positive	2	11	0.001	0.032
14	Mountains Of Creta	Positive	2	Quietly Tuesday	Positive	2	11	0.001	0.032
15	Angel Manor	Positive	2	Kismet North-winds	Positive	2	10	0.003	0.027
16	Angel Manor	Positive	2	Kalepa	Positive	2	9	0.011	0.022
17	Kalepa	Positive	2	Yumix Prada	Positive	2	9	0.011	0.022
18	Quietly Tuesday	Positive	2	Yumix Prada	Positive	2	8	0.036	0.018
19	Angel Manor	Positive	2	Mediterra-neo OC	Uncertain	2	9	0.011	0.022
20	Angel Manor	Positive	2	Port Babbage	Uncertain	2	8	0.036	0.018
21	Angel Manor	Positive	2	Trianwe	Uncertain	1	8	0.036	0.018
22	Agriopis	Uncertain	1	Harshap	Negative	1	15	0.000	0.050
23	Agriopis	Uncertain	1	Prefabrica	Negative	1	8	0.036	0.018
24	Cavettaz	Uncertain	2	Kalepa	Positive	2	15	0.000	0.050
25	Cavettaz	Uncertain	2	Mountains Of Creta	Positive	2	14	0.000	0.046
26	Cavettaz	Uncertain	2	Quietly Tuesday	Positive	2	15	0.000	0.050
27	Cavettaz	Uncertain	2	Yumix Prada	Positive	2	15	0.000	0.050
28	Alice	Uncertain	2	Mountains Of Creta	Positive	2	10	0.003	0.027
29	Mediterra-neo OC	Uncertain	2	Mountains Of Creta	Positive	2	10	0.003	0.027
30	Port Babbage	Uncertain	2	Yumix Prada	Positive	2	10	0.003	0.027
31	Alice	Uncertain	2	Angel Manor	Positive	2	9	0.011	0.022
32	Canis Beach	Uncertain	2	Kalepa	Positive	2	9	0.011	0.022
33	Alice	Uncertain	2	Kismet North-winds	Positive	2	9	0.011	0.022
34	Cavettaz	Uncertain	2	Kismet North-winds	Positive	2	9	0.011	0.022

35	Canis Beach	Uncertain	2	Yumix Prada	Positive	2	9	0.011	0.022
36	Alice	Uncertain	2	Kalepa	Positive	2	8	0.036	0.018
37	Bracket	Uncertain	2	Misthaven	Uncertain	2	14	0.000	0.046
38	Depoz Specialities	Uncertain	2	Trianwe	Uncertain	1	14	0.000	0.046
39	Canis Beach	Uncertain	2	Cavettaz	Uncertain	2	13	0.000	0.041
40	Misthaven	Uncertain	2	Trianwe	Uncertain	1	12	0.000	0.036
41	Depoz Specialities	Uncertain	2	Misthaven	Uncertain	2	9	0.011	0.022
42	Bracket	Uncertain	2	Trianwe	Uncertain	1	9	0.011	0.022
43	Canis Beach	Uncertain	2	Trianwe	Uncertain	1	8	0.036	0.018

Within the significant positive co-occurrences discovered, 58.14% are for VEs sharing the same "uncertainty type" (i.e., positive, negative, or uncertain), and 41.86% are co-occurrences between VEs originally of different types. Regardless of uncertainty, we also re-assessed these percentages in terms of the original *cluster* memberships (specified in Table 7.6 and again in Table 7.8), given that, while MBC does provide uncertainty values for each case, it *does* still assign each case to a single cluster. Once we verified the match between the cluster of VE 1 and VE 2 within each co-occurrence pair (while ignoring information on VE uncertainty), the percentage of agreement soared to 88.37%, with only 11.63% of significant positive co-occurrences taking place between incongruent clusters (positive vs. negative). These co-occurrence results are particularly impressive, since they are based on an MBC classification using PAD data from a different participant sample, and where the Rift was not used (see Chapter 5, p. 227).

This suggests that MBC classification outcomes bear some similarity to how humans themselves classify emotional stimuli - a similarity which may arise from the clustering algorithm and the human classifiers both relying on PAD dimensions as their starting point for the classification (rather than any additional / other dimensions). Also, MBC validity when mimicking how humans classify emotional information is highly relevant for the soundness of algorithmically sampling database stimuli for use in human research: this assumption (for which there is now some empirical evidence) has formed the basis of our work in Chapter 2 (p. 65) and Chapter 3 (p. 99). In other words, if clustering algorithms can closely replicate how humans themselves would classify the stimuli, then using such categories of stimuli as independent variables in research would be less prone to error.

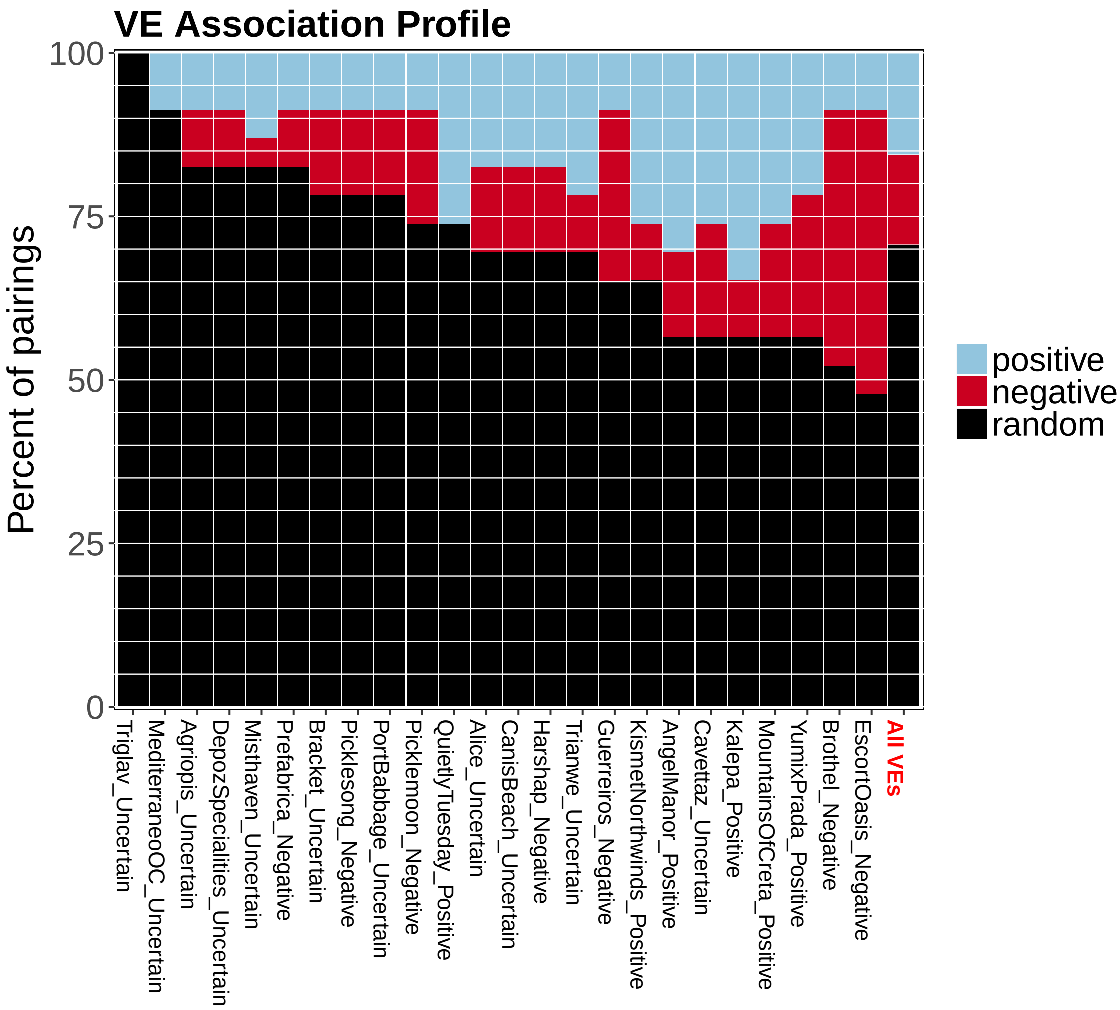


Figure 7.8: VE pair profile, including a representation for the overall proportion of positive, negative and random co-occurrences which the VEs were involved in.

7.3.6 Hierarchical clustering

Finally, we investigated whether one additional form of cluster analysis also provided meaningful groups of VEs, based on the binary, presence/absence matrix. We used hierarchical cluster analysis, as this presents a suitable option for calculating binary distances in `R`[9]. Based on this clustering algorithm and distance calculation method used, Figure 7.10 (p. 317) represents the distances at which various clusters merge into ever larger groups (i.e., the tree height on the y axis). Cases and clusters are shown to be quite distinct and only merge at larger tree heights in the dendrogram, however when they do, they still show considerable similarity with the originally intended VE

[9] From the `?dist` help file in `R`: "The vectors are regarded as binary bits, so non-zero elements are 'on' and zero elements are 'off'. The distance is the proportion of bits in which only one is on amongst those in which at least one is on."

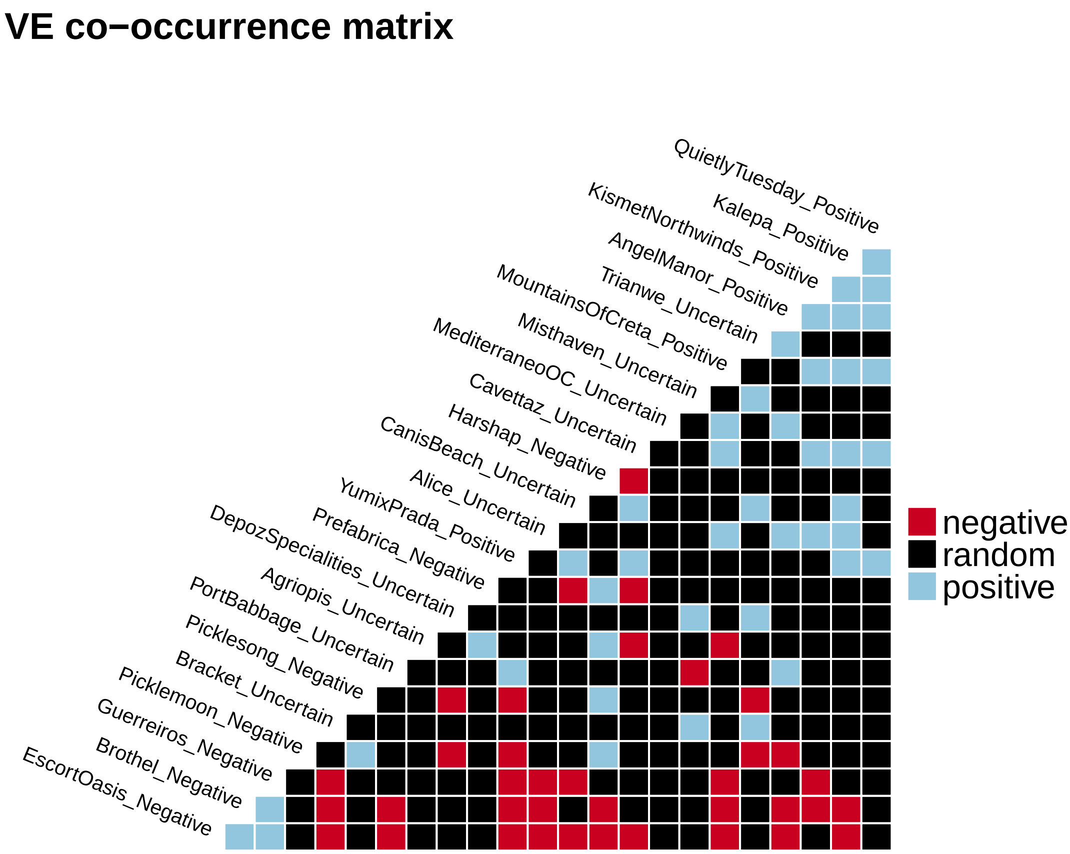


Figure 7.9: VE co-occurrence matrix. Interestingly, there are two "nodes" of positive and negative VEs at the opposite corners of the diagram. On the one hand, the positive, interconnected VE pairs include `Quietly Tuesday`, `Kalepa`, `Kismet Northwinds` and `Angel Manor`, whereas the negative interconnected co-occurrences include the VEs: `Guerreiros`, `Brothel` and `Escort Oasis`.

types (i.e., positive, negative and uncertain). In fact, computing a Rand index of overlap between this hierarchical clustering and the intended VE structure leads to a sizeable value of 0.714.

7.3.7 Comparing model-based cluster analysis results with human performance

Since participants in this study also provided the typical PAD SAM ratings for all stimuli, we also tried to replicate the bipolar classification structure found previously for VEs (see Section 5.3.5, p. 253). Indeed, previous findings ($k = 2$) were upheld for this new data, with the cluster structure and uncertainties detailed in Figure 7.11 (p. 320). The top three models were based on BIC values were: EEV (for $k = 2$, and BIC = -130.993), EEE ($k = 1$ and BIC = -133.252), and finally another EEV model (for $k = 1$, and BIC = -133.252). For the optimal model of these three (i.e., $k = 2$), the newly-generated

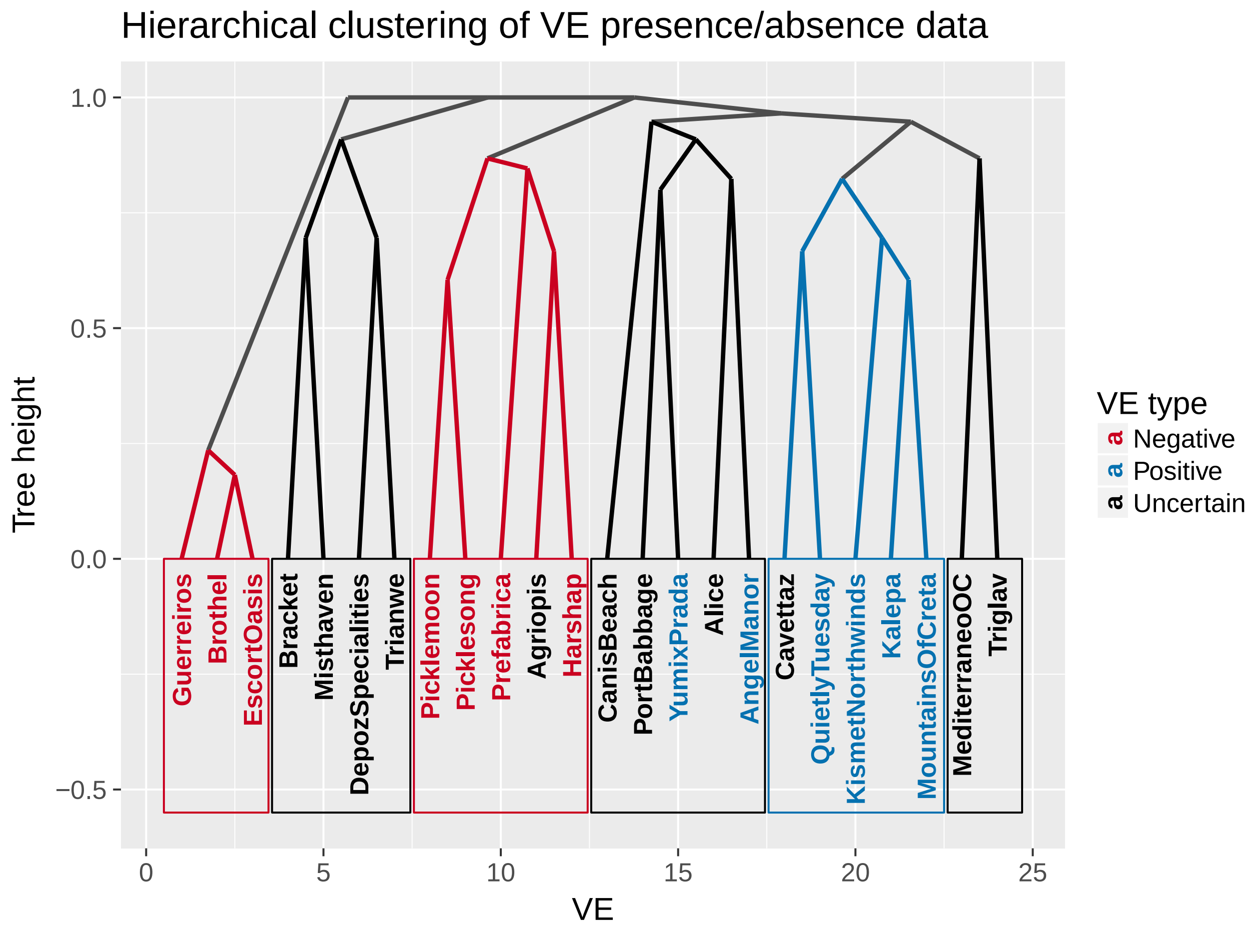


Figure 7.10: VE dendrogram, with colour-coding of original / intended VE types. The tree branches and the rectangles around group labels have borrowed the colour of the most frequently occurring VE type underneath.

uncertainties and cluster memberships (based on ratings participants provided within this study, rather than a previous one) are specified in Table 7.9 below:

Table 7.9: VE classification: MBC solution

No	VE	Average Valence	Average Arousal	Average Dominance	Uncertainty (current study)	Classification (current study)
1	Picklemoon	2.000	6.167	2.067	0.000000	1
2	Picklesong	3.533	5.133	2.933	0.000000	1
3	Guerreiros	3.500	4.767	2.933	0.000000	1
4	Escort Oasis	3.433	4.667	3.033	0.000000	1
5	Agriopis	2.733	4.567	3.200	0.000001	1
6	Brothel	2.767	4.533	3.133	0.000001	1
7	Harshap	3.133	4.467	3.133	0.000007	1
8	Triglav	4.067	4.533	3.600	0.000019	1
9	Prefabrica	3.333	3.800	3.267	0.026261	1
10	Port Babbage	4.167	3.700	3.867	0.352894	1
11	Canis Beach	5.900	3.033	5.033	0.000000	2

12	Yumix Prada	6.100	2.967	4.733	0.000001	2
13	Cavettaz	6.033	2.367	5.100	0.000001	2
14	Misthaven	5.500	3.433	4.733	0.000005	2
15	Mediterraneo OC	5.033	2.800	3.767	0.000010	2
16	Kalepa	5.500	3.000	4.567	0.000029	2
17	Quietly Tuesday	5.267	2.833	4.100	0.000059	2
18	Mountains of Creta	5.433	2.600	4.467	0.000064	2
19	Bracket	5.367	3.033	4.467	0.000097	2
20	Trianwe	4.967	3.100	4.833	0.000282	2
21	Depoz Specialities	5.267	3.100	4.167	0.000333	2
22	Kismet Northwinds	4.867	3.133	3.933	0.001712	2
23	Alice	4.633	3.167	3.767	0.003019	2
24	Angel Manor	4.433	2.967	3.867	0.010725	2

By updating the linear mixed model previously summarised in Table 7.7 with the newly-computed uncertainties from the current MBC solution, previous findings were replicated - see Table 7.10. As usual, all continuous predictors were scaled. Random intercepts were added for both participants and SLURLs, with the former allowed to vary by intended VE category (positive, negative or uncertain).

Table 7.10: VE uncertainty model updated after newly-generated MBC solution

	Model 1
(Intercept)	4.29^{***}
	(0.32)
Uncertainty score (current study)	2.46^{*}
	(1.11)
Intended VE category: Positive	0.57
	(0.69)
Intended VE category: Uncertain	-0.05
	(0.31)
Uncertainty score (current study) × Intended category: Positive	-3.72
	(3.02)
Uncertainty score (current study) × Intended category: Uncertain	-2.49^{*}
	(1.11)
AIC	2672.09
BIC	2736.20
Log Likelihood	-1322.05
Num. obs.	720
Num. groups: Participant	30
Num. groups: SLURL	24
Var: Participant (Intercept)	1.18
Var: Participant × Group label: Positive	1.72
Var: Participant × Group label: Uncertain	0.68
Cov: Participant (Intercept) × Group label: Positive	-0.44
Cov: Participant (Intercept) × Group label: Uncertain	-0.16

Cov: Participant × Group label: Positive × Group label: Uncertain	1.01
Var: SLURL (Intercept)	0.08
Var: Residual	1.89

$^{***}p < 0.001$, $^{**}p < 0.01$, $^{*}p < 0.05$

7.3.8 Multinomial models predicting individual classification structures

To gain a deeper understanding of the factors that may be shaping participant-made classifications, we assessed the amount of overlap between the classification created by each participant for the 24 VEs, and the MBC classification generated from their PAD ratings and described in the previous subsection. For this purpose, χ^2 tests were carried out for each participant, with p values and Cramer's φ coefficients extracted for inspection (see Figure 7.12, p. 321). An example cross-tabulation between each person's manual grouping and the MBC VE groupings is also present in Table 7.11:

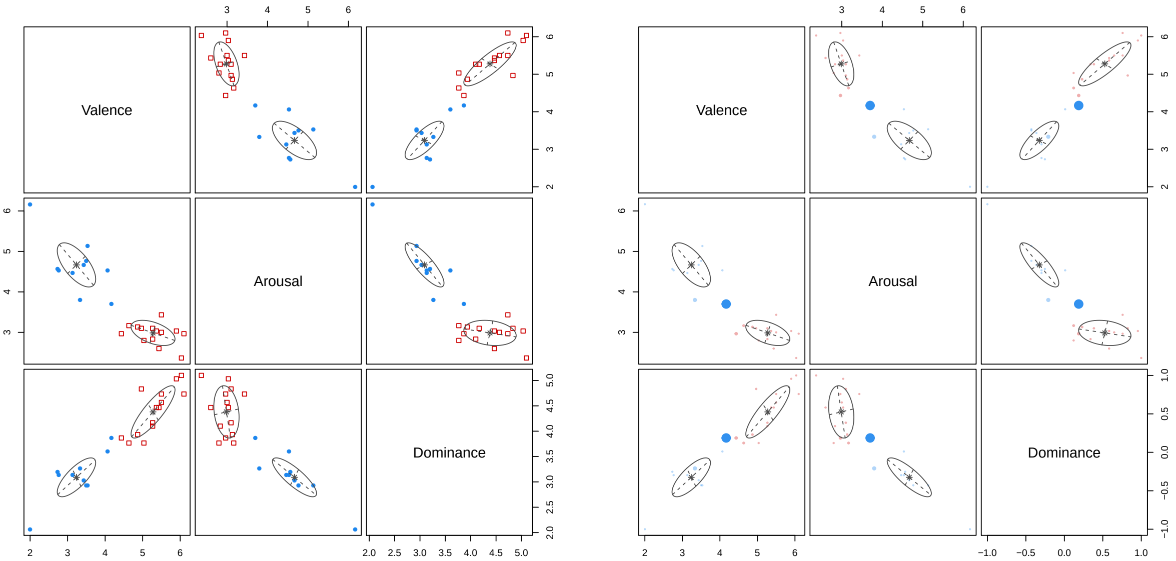
Table 7.11: Example of a manual vs automatic cross-tabulation for one of the participants.

User category	Cluster 1	Cluster 2
bizzare	3	2
disgust	2	0
dislike	0	2
enjoyable	4	2
unclassifiable	1	8

Even if in most cases, there was a significant association (sometimes with high effect sizes) between automatic and human classifications, we wanted to further test the relative contributions of several crucial predictors on participant categories. Because categories differed from participant to participant, and because no *multinomial* mixed effects models are easily available for the statistical software of choice (`R`), we instead created one multinomial model for each participant in order to predict the classifications they produced, from: the sample-wide average Valence, Arousal and Dominance scores, as well as the MBC classification and uncertainties achieved by the VEs.

Of the total sample of 30 participants, 16 of them chose not to use the 'unclassifiable' category, whereas the remaining 14 did use this pre-existing category. For the former, multinomial models were created separately with the 'unclassifiable' category as the baseline, whereas the other participants had the baseline category set to whichever VE they created and was also the largest of all. The results from these models are included in detail - with three pseudo R^2 measures - in Appendix section F.3 (p. 496), and are also present in Figures 7.13 (p. 323) and 7.14 (p. 323) in summary form.

In terms of by-coefficient effect sizes (for the significant predictors only), Wald tests were used to compare the intensity / the usefulness of each predictor, across participants.



(a) Stimulus classification for VEs into the 2 clusters.

(b) VE membership uncertainty: the larger and lighter the points, the higher the uncertainty.

Figure 7.11: VE clusters created using model-based clustering, which replicate previous results by reconfirming a $k = 2$ solution.

This was possible due to Wald z values representing standardised (therefore, comparable) versions of the raw regression coefficients. In cases where the Wald z value was larger than 10 (14.57% of cases), this was taken as a sign of problems in the estimation (e.g., modelling noise instead of signal for a small N etc.), and such values were removed. The remaining values are plotted using boxplots and represent how strongly VE categories (pooled across participants) differed from the baseline category - where the highest median value belongs to the Arousal predictor (see Figure 7.15, p. 324). Hence out of all predictors considered, Arousal was the most important influence on human classifications of emotional information.

7.3.9 Relative and absolute VE classifiability

The approach above (i.e., of creating one multinomial model per participant) may have been underpowered due to multiple predictors being used with only $N = 24$ VEs in each case. This could have obscured / underplayed various relationships between variables. Due to this, we re-structured the data (and the outcome variable) to represent, for each VE, the number of participants who managed to classify it into whichever category (as opposed to placing it within the given 'unclassifiable' category). This continuous outcome, which we named 'relative classifiability', was then predicted by single variables, in order to avoid the previous power-related concerns. The results of these linear mod-

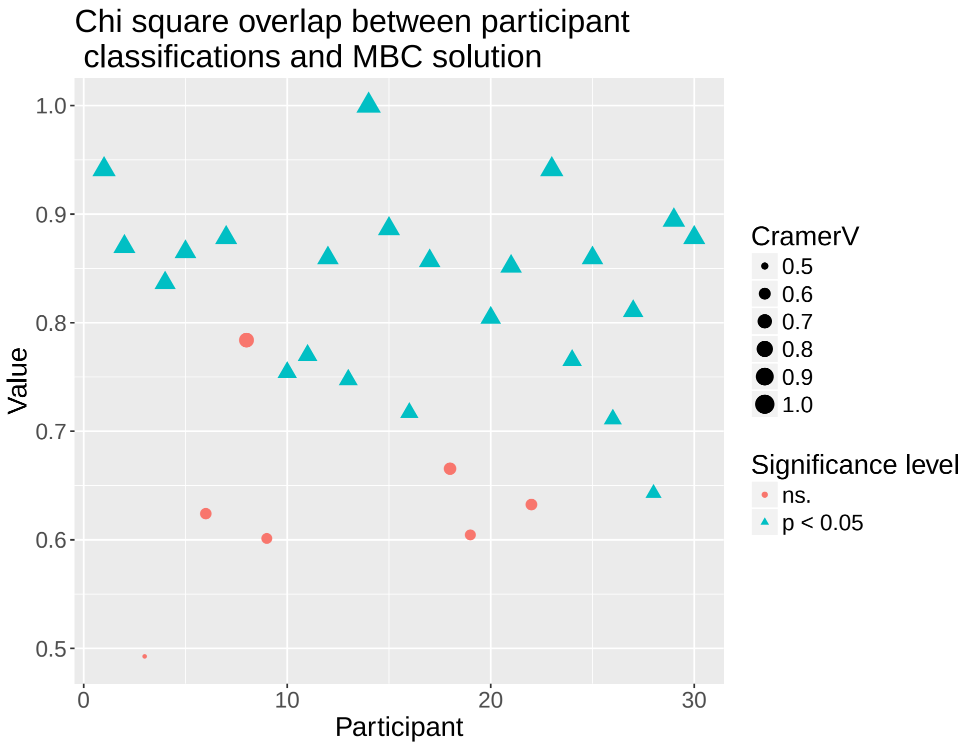


Figure 7.12: VE cross-tabulations. The p values and Cramer's φ values extracted from by-participant χ^2 tests often show a high level of association between the MBC classification and participants' own.

els are presented in Table 7.12 (p. 321). As is customary with all previous models, continuous predictors were scaled.

Table 7.12: Relative VE classifiability models. The first column includes the measures considered for predicting our measure of 'relative classifiability'. The column headings represent the labels attributed to the models, e.g., 'Full Predictors' is an (underpowered) model including all available measures for predicting our outcome, whereas 'Valence' is the label for a simple linear regression model using just Valence scores to predict classifiability. Such simple regression models may better allow to detect effects within this small sample.

	Full Predictors	Valence	Arousal	Dominance	MBC Uncertainties	MBC Classification
(Intercept)	29.60^{***}	28.75^{***}	28.75^{***}	28.75^{***}	28.75^{***}	29.30^{***}
	(0.81)	(0.25)	(0.24)	(0.25)	(0.25)	(0.38)
Valence	0.34	−0.45				
	(0.88)	(0.25)				
Arousal	0.22		0.54^{*}			
	(0.66)		(0.24)			
Dominance	0.01			−0.45		
	(0.74)			(0.25)		
Uncertainty	−0.55				−0.38	

	(0.29)				(0.26)	
MBC Classification: Cluster 2	−1.45					−0.94
	(1.32)					(0.49)
R^2	0.32	0.13	0.19	0.13	0.09	0.14
Adj. R^2	0.14	0.09	0.15	0.09	0.05	0.10
Num. obs.	24	24	24	24	24	24
RMSE	1.17	1.20	1.16	1.20	1.23	1.19

$^{***}p < 0.001$, $^{**}p < 0.01$, $^{*}p < 0.05$

We further created an alternative type of model where we measured '*absolute* classifiability' as a binary outcome for each VE: in other words, whether or not any of the participants in the sample had placed a given VE within the 'unclassifiable' category at any point. Results for these logistic models are presented in Table 7.13 (p. 324).

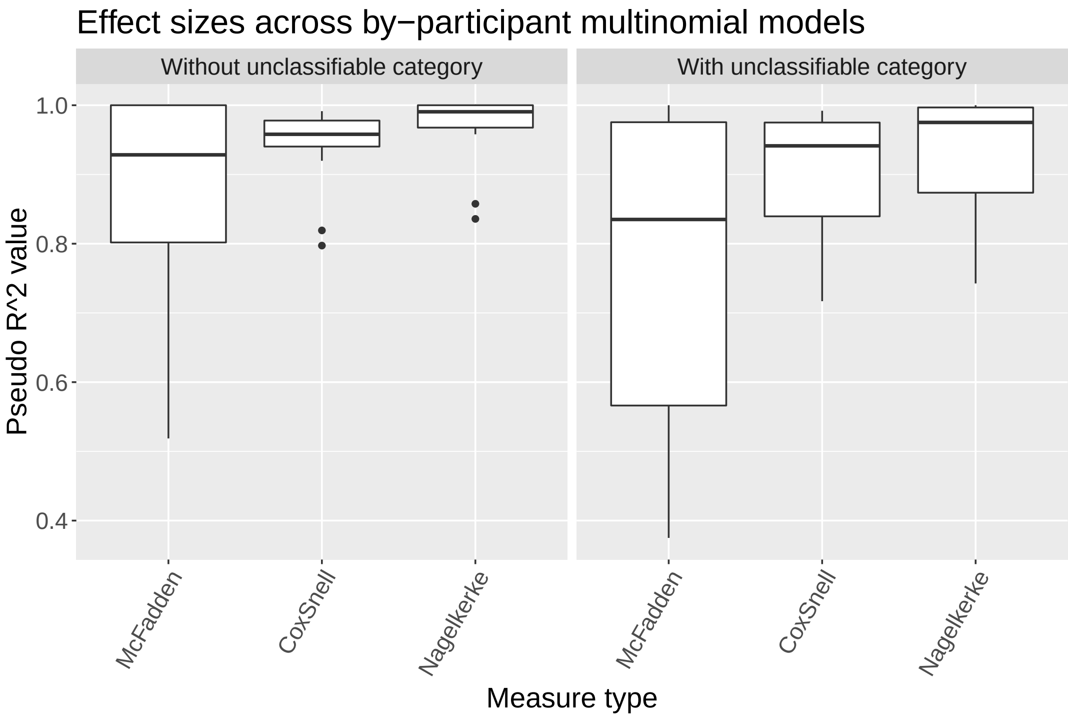


Figure 7.13: VE by-participant multinomial models: pseudo R^2 measures usually showing good model fit.

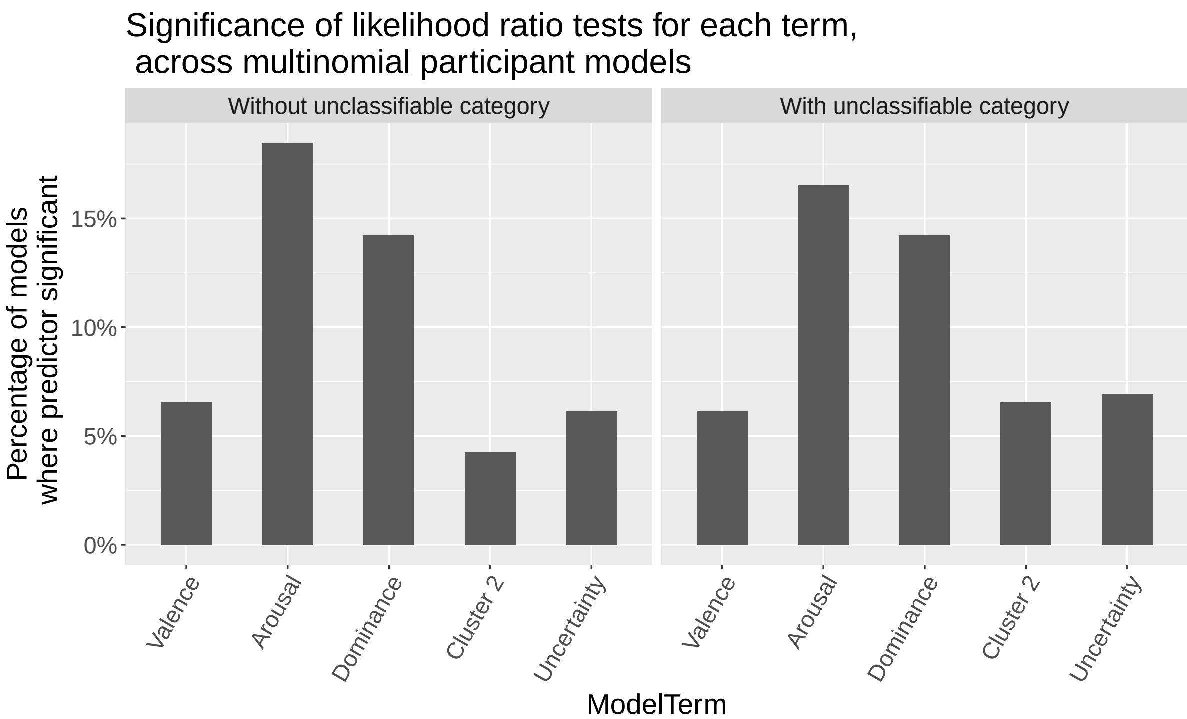


Figure 7.14: VE by-participant multinomial models: significance of model terms. Arousal emerges as the dimension which is most often useful for predicting how participants classify emotional information.

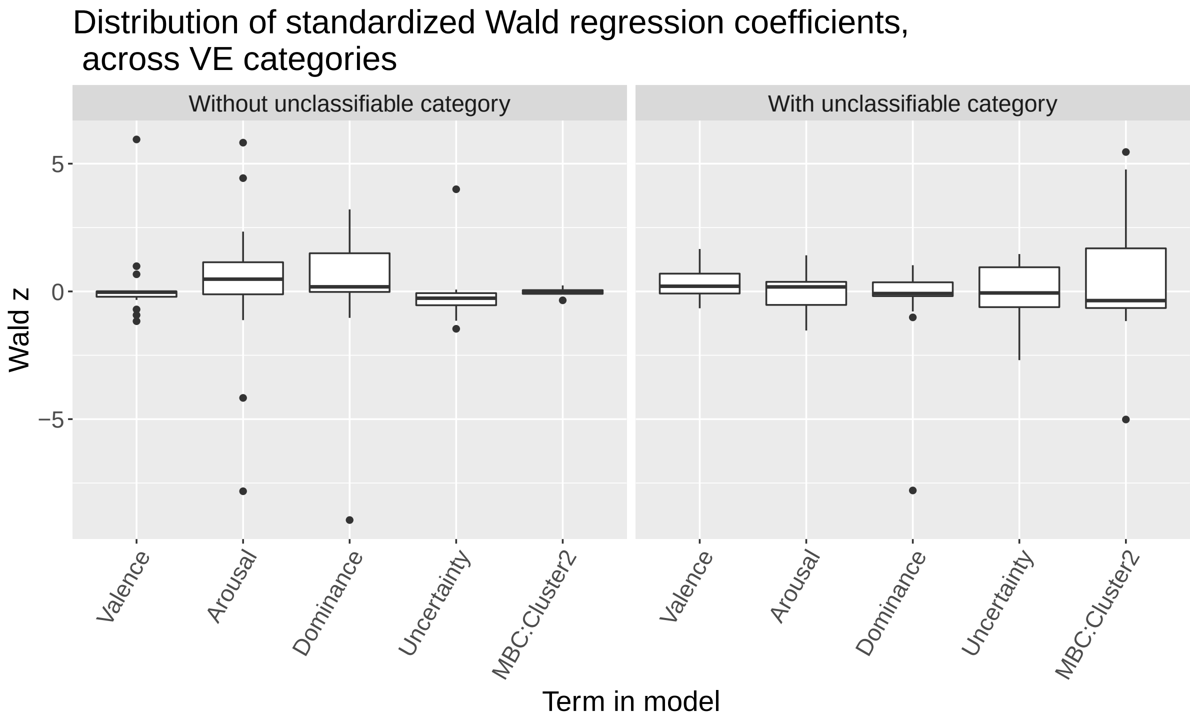


Figure 7.15: VE by-participant multinomial models: Wald z coefficients. The median Wald z score reconfirms the importance of Arousal in influencing participant classifications.

Table 7.13: Absolute classifiability models.

	Full Predictors	Valence	Arousal	Dominance	MBC Uncertainties	MBC Classification
(Intercept)	-199.74	-1.09^{*}	-1.17^{*}	-1.16^{*}	-314.96	-0.00
	(206.01)	(0.54)	(0.58)	(0.56)	(328.15)	(0.63)
Valence	-2.24	-1.10^{*}				
	(2.75)	(0.55)				
Arousal	6.38		1.46^{*}			
	(4.30)		(0.63)			
Dominance	-1.50			-1.27^{*}		
	(2.07)			(0.63)		
Uncertainty	-828.22				-1373.42	
	(895.59)				(1432.39)	
MBC Classification: Cluster 2	16.43					-1.79
	(10.99)					(0.99)
AIC	25.94	28.09	25.21	27.30	26.20	29.35
BIC	33.01	30.44	27.57	29.65	28.55	31.70
Log Likelihood	-6.97	-12.04	-10.61	-11.65	-11.10	-12.67
Deviance	13.94	24.09	21.21	23.30	22.20	25.35
Num. obs.	24	24	24	24	24	24

$^{***}p < 0.001$, $^{**}p < 0.01$, $^{*}p < 0.05$

7.3.10 Participant-made hierarchies within VE groups

A final model was generated in order to investigate the hierarchies of VEs which the participants created within each category. Namely, since participants were asked to rank their VEs within each group they produced, we wanted to investigate whether these rankings were related to MBC uncertainties, as well as other measures - see Listing 7.2. However, we ignored the individual categories created by participants, in themselves. In this case, only the extent to which a VE was a good representative of its category was the point of interest, and not which specific category that is.

Listing 7.2: `R` code snippet: Predicting VE hierarchies using a mixed-effects model. Any continuous predictors were scaled. Random terms included random intercepts by VE (i.e., `card label`), as well as by participant (i.e., `login entry`).

```
# lmer() model structure:
s <- function(x){ scale( x, center = TRUE, scale = TRUE ) }
hierarchy_mod <- lme4::lmer( `sorted position` ~
                               s( UncertaintyCurrentStudy ) +
                               as.factor( ClassificationCurrentStudy ) +
                               s( Valence ) +
                               s( Arousal ) +
                               s( Dominance ) +
                               ( 1 | `login entry`) +
                               ( 1 | `card label` ),
                             data = dat )
```

Not only did Arousal frequently *and* quite intensely affect participant classification patterns at an individual level (see Section 7.3.8, p. 319), but it also influenced the way that VEs were ordered hierarchically within a category - see Table 7.14 below: the better the representative (i.e., the smaller the rank within a category), the higher the associated Arousal ratings for that VE. Interestingly, in this case, it was shown that MBC uncertainties did not correspond directly to these participant-made measures of uncertainty (i.e., MBC uncertainty was a non-significant predictor of within-category, human-made hierarchies).

Table 7.14: VE participant-made classification hierarchies / uncertainties, predicted by Arousal scores:

Coefficients	Predicting hierarchies / ‘participant-made uncertainty scores
(Intercept)	2.68^{***}
	(0.27)
Uncertainty (current study)	0.04
	(0.09)
MBC Classification (current study): Cluster 2	0.06
	(0.42)
Valence	-0.38

	(0.27)
Arousal	-0.77^{***}
	(0.20)
Dominance	-0.21
	(0.23)
AIC	2783.51
BIC	2824.72
Log Likelihood	-1382.75
Num. obs.	720
Num. groups: login entry	30
Num. groups: card label / VE / SLURL	24
Var: login entry (Intercept)	0.25
Var: card label / VE / SLURL (Intercept)	0.05
Var: Residual	2.53

$^{***}p < 0.001$, $^{**}p < 0.01$, $^{*}p < 0.05$

7.3.11 Category verbal labels

To further verify how meaningfully participants completed the sorting task, as well as how valid the distinction between clearly negative, clearly positive and uncertain items truly was, we created wordclouds to show the most frequently used verbal labels used by participants, depending on the type of VE they referred to. The labels were lemmatised using TreeTagger software (Schmid, 2013) and the associated `koRpus R` package, and then stemmed using the `SnowballC R` package, in an attempt to standardise the word forms used by participants. Results are shown in Figure 7.16 (p. 327), and suggest that for positive and negative VEs, the most frequently occurring words tend to match the intended Valence of the VEs. As might be expected, the pattern is more mixed for the uncertain VEs.

7.4 Discussion

In this study, participants were asked to **navigate** through 25[10] VEs using an HMD, **rate** them on the PAD model, and also **classify** them freely according to the emotional states they conveyed. The 25 stimuli had been selected via a data-driven strategy based on *previous* research and data. Briefly, a larger set of 51 VEs (see Chapter 5, p. 227) was successfully grouped into 2 clusters (i.e., one generally positive, and one generally negative cluster, with VEs varying in terms of their membership uncertainty). From these data, we selected the subset of 25 VEs as stimuli for the current study (see Section 7.3.2.2, p. 302). These included: the 7 best representatives from each of the two

[10] This ultimately turned into 24 VEs, as one VE went offline during data collection and was excluded from the analysis.

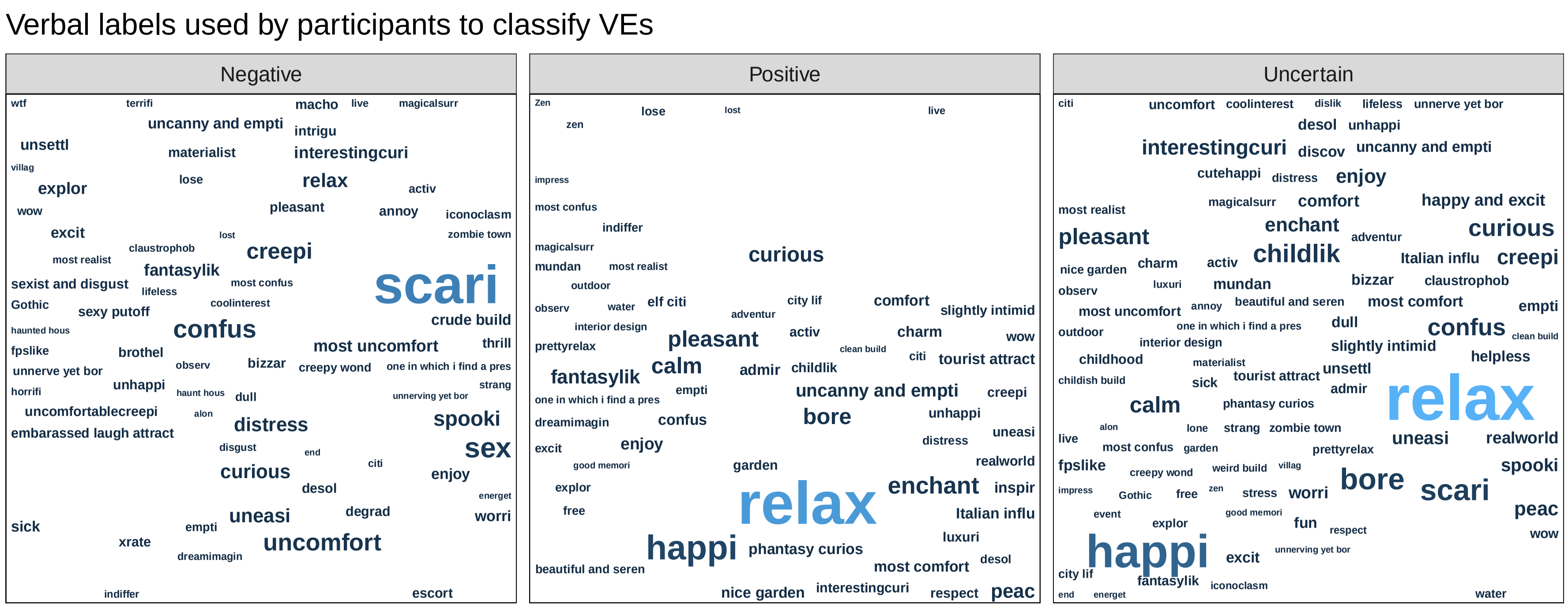


Figure 7.16: VE wordclouds, split by VE type (positive, negative, or ‘uncertain’. Word size is proportional to frequency within the (small) corpus of participant-made category labels. For positive and negative VEs, the most frequently occurring words tend to match the intended Valence of the VEs. As might be expected, the pattern is more mixed for the uncertain VEs.

clusters (i.e., a 'clearly' positive and 'clearly' negative VEs), and then, 11 further items varying in uncertainty from the space between these two clusters (i.e., 'uncertain' VEs).

7.4.1 General classification tendencies

Firstly, using the new data collected in this study, we replicated the $k = 2$ cluster structure for the 24 VEs. Furthermore, almost all VEs used here as stimuli maintained their dichotomous cluster membership from the previous study, regardless of uncertainty. Hence, two participant samples rated these VEs similarly enough on the PAD model for the MBC algorithm to detect and replicate largely the same clustering structure.

In terms of participants' own free classifications of the same 24 VEs, they were grouped on average into ≈ 7 groups, but the number of groups each participant opted for was influenced by VE uncertainty (as computed using MBC): the further into the space between the positive and negative clusters a VE was situated, the higher the chance that it would be placed into one of multiple fine-grained VE groups. In contrast, if a VE was 'clearly' negative or 'clearly' positive (i.e., presented low uncertainty), then participants were more likely to place it in one of fewer, larger and coarser categories. This hints at a **first link** between MBC and human classification of virtual, emotional stimuli: participants may have been unsure of the "common thread" between the uncertain VEs, and hence felt the need to create more specific, fine-grained categories for them. In the case of the "clear-cut" positive or negative VEs, this may have been seen as unnecessary, leading to VEs being grouped into broader classes.

In fact, if pooling the classification data across the entire sample, hierarchical clustering identified 6 groups of VEs, which largely mapped onto the MBC clustering structure - with 'positive', 'negative' and 'uncertain' VEs tending to be classified together and not mixed. This is a good indication that, pending further research, clustering algorithms may serve as a first approximation of an algorithm-level model of emotional information processing.

7.4.2 Co-occurrence patterns

Within participant classifications, we assessed how often any given pair of VEs found themselves sharing the same participant-made category, regardless how large the category was, or how many other categories existed. We found that $\approx 30\%$ of VE pairs / co-occurrences were non-random (i.e., occurred either more, or less frequently than chance level). Most noticeably, there were two "nodes" of particularly interconnected pairs (see Figure 7.9, p. 316). The positive node included the VEs: `Quietly Tuesday`, `Kalepa`, `Kismet Northwinds` and `Angel Manor`. Interestingly, all of these VEs presented almost null uncertainties of belonging to the generally positive MBC cluster. The negative node included the VEs: `Guerreiros`, `Brothel` and `Escort Oasis`. These too presented very low uncertainties (i.e., below 0.001).

This suggests a **second link** between how MBC and humans treated this data: participants may be detecting the clear representatives of a cluster similarly to MBC, and categorising them "correctly" according to both their cluster membership (positive / negative), as well as their uncertainty - with very certain items showing strong inter-relationships and frequently co-occurring within the same participant-made categories. Almost all participants in the sample grouped together the three negative VEs into the same category ($N = 26$ or 27, depending on the co-occurring pair in this node, see Table 7.8, p. 312), whereas for the 4 positive VEs, on average 11.5 participants (i.e., between 9 and 14) placed these pairs together in the same category.

The 3 negative VEs very often grouped together also happened to be highly similar semantically (i.e., they all contain erotic content), which might explain why the positive pairs presented relatively lower co-occurrence frequencies (as similarities between positive VEs tended to be mostly emotional, rather than semantic as well). This also suggests that participants were using one or more additional dimensions in their classification (such as semantic themes etc.), as opposed to MBC, which relied strictly on Valence, Arousal, and Dominance ratings. In addition, other VEs also showed similarly low uncertainties within the positive and negative MBC clusters, and yet were not part of these co-occurrence nodes. This too is an indication that participants may be relying on extra dimensions and rules for completing this classification task.

However, regardless of whether or not a VE appeared within a nexus of co-occurrences, generally speaking across all the individual positive co-occurrences which were flagged as significant, 88% of pairs were established between members of the same MBC cluster. This constitutes a **third link** between MBC results and the manner in which human classifiers solved this task. This similarity between MBC and how humans classify emotional stimuli may signal that both the algorithm and human classifiers are relying on PAD dimensions for their classification. Also, MBC validity when mimicking how humans classify emotional information is important when considering strategies for algorithmically sampling stimuli to be used in human research: this assumption has formed the basis of our work in Chapter 2 and Chapter 3. If clustering algorithms can closely replicate how humans themselves would classify the stimuli, then research involving distinct categories of stimuli as independent variables would be less prone to error and thus appear as more coherent to human participants.

When revising this calculation to incorporate uncertainty information, i.e., separately assessing co-occurrences between the "clearly" positive, "clearly" negative and "uncertain" VEs, then the percentage of co-occurrences which still occur between congruent VE types is lower: 58%. Finally, it is very likely that participants ignore parsimony-related constraints when creating a classification - unlike MBC. Indeed, the number of categories created by humans almost always far outstrips the parsimonious $k = 2$ solution provided by MBC. In addition to the possibility that that participants may also use extra dimensions in addition to PAD when classifying the stimuli, these findings suggest that future research should either investigate what additional dimensions may be useful to include in order to bridge the gap between MBC and human classification, or assess additional clustering algorithms, or both.

7.4.3 Predicting individual classification patterns

We investigated the classification process from two perspectives: firstly from the participants' point of view - in terms of how they categorised VEs and ranked them in order of prototypicality, and secondly, from the VEs' point of view - in terms of their relative[11] or absolute "classifiability"[12]. When carrying out the classification task, participants were offered an optional "Unclassifiable" category, in case any VEs remained which did not fit with their classification structure. Nearly half of the participants (14 out of 30) resorted to using this category.

Firstly, we were able to predict the actual groupings of stimuli created by participants in a series of multinomial models, one for each person - with the baseline set to either the "Unclassifiable" category if this was used by a given participant, or if this category was left unpopulated, participants' most inclusive (largest) category served as baseline. Thus, the predictors considered (i.e., the average Valence, Arousal, Dominance scores achieved by each VE, their MBC cluster membership, and their uncertainty values) were all tested in terms of their contribution towards increasing the likelihood that an item was placed into a more "specific" category than either the unclassifiable one, or than the most generic one.

Of the predictors tested across the 30 models, Arousal was significant most often, and presented a higher median effect size compared to all other predictors - regardless of whether the baseline category was the "Unclassifiable" or the largest one. Dominance followed in terms of its usefulness across these by-participant models. This is further evidence that *all* PAD dimensions should be considered in research, rather than just Valence (see our discussion from Part II, p. 63). In fact in this case, Valence as well as the MBC classification and the associated uncertainties were all considerably less useful in these models compared to Arousal and Dominance. These findings should not however be taken as equally applicable to all participants: considerable variety

[11] As a reminder, relative classifiability refers to the number of times a given VE was placed into the "Unclassifiable" category by participants, vs. was successfully included in some other category.

[12] This is a binary outcome measuring whether or not a VE has ever been considered "Unclassifiable", by any participant.

was present across which (set of) predictor(s) proved useful for a given participant. This opens up the interesting possibility that individual differences may consistently be shaping how participants classify emotionally-toned information.

In terms VE rankings within participant-made categories, Arousal *also* had a role to play: the higher the Arousal score associated with a VE, the more likely it was for this VE to be regarded as "prototypical" or very representative for the category in which it had been placed. Interestingly, MBC uncertainties, however, did not bear any relationship to how human participants sorted the VEs within their categories. This suggests that while uncertainties do influence category size (as discussed above), they do not directly correspond to how participants identify representative category elements.

Revisiting the original point regarding the influence of Arousal - memory effects may be mediating the relationship between this dimension, and stimulus groupings and hierarchies. Arousal has been found to boost memory for emotional stimuli (Cahill & McGaugh, 1995; Kensinger & Corkin, 2004; LaBar & Phelps, 1998), so that here, VEs which are more memorable may be driving the classification process itself, and may also be positioned as more representative of the categories they are part of. In this way, when faced with the classification task after visiting 24 VEs, participants may be using the most arousing VEs as "anchors" around which to organise their classification process. If so, this would explain why Arousal is tied to both the category member prototypicality, as well as classification structure (i.e., how VEs are assigned to groups). In other words, participants may be *initialising* categories with the more arousing VEs positioned as category prototypes (i.e., the most representative items in a category). This could be investigated further alongside the idea of "category drift" using the desktop recording data also collected in this study. Time constraints did not allow this investigation at present, but it could be carried out in the future.

Secondly, and because of sample size concerns, we also investigated how "classifiable" the 24 VEs were, by predicting both relative classifiability (via simple linear regression) and absolute classifiability (via logistic regression). For the former, Arousal yet again emerged as the only significant predictor, thus conceptually replicating previous findings. For the latter, logistic models flagged all 3 PAD dimensions as (individually) significant predictors for whether or not a VE had ever been placed into the "Unclassifiable" category, by any participant. Interestingly, the more negative a VE was (or the lower the Valence) on average, the more likely it was for the VE to be classed under some category, rather than be left unclassified. Independently, Dominance behaved in the same way: the less Dominance-inducing the VE, the higher the chance of it being successfully classified. This similarity in results is likely due to the extremely high correlation between average Valence and average Dominance scores in this dataset: $r = 0.939$. In terms of Arousal, the findings are consistent with all those presented above: the more arousing a

VE was, the easier it was to classify.

7.4.4 Limitations

The work presented here might be improved upon in future research in three main areas: conceptual, methodological, and statistical: on a conceptual level, the experimental tasks probably engaged some degree of imagination on the part of participants, given they were asked to rate how they would feel *if* they found themselves in the situations illustrated by the VEs. This way of formulating the task was adopted to avoid that participants passively rate or classify only the stimuli themselves, instead of the emotional states they conveyed. However, in this fashion it becomes difficult to distinguish between "raw" emotional reactions and an emotionally-toned imaginative process. Furthermore, it is likely that memory processes were also closely associated with the experimental tasks, as the PAD rating procedure occurring after the exploration of the 25 VEs did not offer a reminder menu, and even when this was present in the case of the classification task, it is likely that participants still relied heavily on memory, and did not consult the reminder menu for each and every VE before classifying it. Hence, our data should be seen as a result of the interplay between various psychological processes, rather than "raw emotionality" alone.

In terms of methods-related limitations, several improvements could be further introduced: the small sample of participants could be increased and sampled from one single cultural background, instead of a variety. In addition, more advanced technical equipment (e.g., currently the HTC Vive) could be used instead of the Oculus Rift DK1, which is now obsolete and suffers from the same limitations discussed in Section 6.4.1 (p. 285), e.g., low resolution and time lag etc. Finally, measures could be introduced in order to make sure that participants follow instructions carefully in this complex design, given that some of them appear to have occasionally grouped VE by semantic themes / visual content, than purely by the emotional content - which is implied by some of the category labels they provided (e.g., "`Village`", as opposed to "`creepy`").

Finally, in terms of statistical limitations, it would have been ideal to take into account the *nested structure* of the co-occurrence data (VEs within categories within participants). This was not possible here, but should be aimed for in the future in order to better detect individual differences, which were shown to affect predictive models of how the VEs were classified.

7.4.5 Future directions

In the previous study comparing PAD ratings between a large monitor and the Oculus Rift HMD, we suggested that the unexpected variation between stimuli might be due to the extremely small sample of VEs tested using both the monitor and the Rift ($N =$

6). In this study, we tested additional VEs with the Rift ($N = 24$, after excluding one inconsistent VE) and measured their PAD properties. Therefore, in the future this new data could be compared to previously gathered data using a monitor only, for the same 24 VEs.

In addition, current results could be replicated with another stimulus type (e.g., films, for which data was already collected using the same experimental paradigm), to reassess the role of Arousal in the classification itself, in predicting 'classifiability', *and* in establishing the hierarchy of items within groups. Testing additional stimulus types would also be useful for checking whether the relationship between co-occurrences and MBC classification is maintained, i.e., if co-occurrences are mainly established between members of the same MBC cluster.

Another line of enquiry arising from this study relates to "category drift", or how participant categories are built and change over time, before the participant believes they have reached the final iteration of their classification solution. This process could be investigated using the desktop recording data that has already been collected - which could simultaneously be used to verify whether the classification process is indeed initialised with very arousing VEs as the most representative.

Pending further research, clustering algorithms may be able to eventually recreate the manner in which humans classify emotional information. For this to be possible, it would be useful to investigate what additional dimensions should be taken into account (above and beyond PAD) in order to bridge the gap between MBC and human classification. Also, additional clustering algorithms could also be tested in case they offer more similar classifications to those produced by human participants.

Finally, individual differences appear to consistently be shaping how participants classify emotionally-toned information, because models predicting individual classifications tended to vary. Therefore, future research could assess which individual characteristics may influence classification patterns. This could be important for, e.g., removing this variation as error from other analyses.

7.5 Conclusions

Interesting links were confirmed between MBC and human classifications of emotional, virtual stimuli. However, these were found to be indirect: for instance, MBC uncertainties for VEs inhabiting the space between the generally positive and negative clusters led to category fragmentation (i.e., these VEs were placed into smaller, fine-grained categories), but were not related to the prototypicality hierarchies within human-made categories. More research will be needed to further investigate these effects, as well as the influence of Arousal on the classification of emotional content - be it VEs, films or other types of stimuli.

Part V

Experience-sampling study, comparing the lab-stimuli with real-life events

Chapter 8

Study 5: Validating lab emotion elicitation methods against affective states occurring in daily life

8.1 Introduction

THE previous work outlined in this thesis has involved affective stimuli of various types: images, words, sounds, film clips and virtual environments (VEs - where the latter have been tested both with and without a Head-Mounted Display). This research attempts to explore how various stimulus modalities should be sampled and matched, and whether there are systematic differences between them.

Our previous findings include that in the case of VEs, PAD ratings are range-restricted relative to the other stimuli, so much so that matching them to other stimulus types becomes very difficult. In addition, the VE clustering structure is reduced to $k = 2$ clusters compared to $k = 4$ or 5 in the case of the other stimuli, suggesting that VEs are perceived more homogeneously, and/or with less nuance, than other types of stimuli used for emotion elicitation. We further discovered that emotional sounds tended to show a level of Arousal which was overall higher than for affective images or words. Such differences suggest that the choice of emotional stimuli may significantly and potentially systematically skew research results, so an informed decision is required concerning which stimulus type to opt for at the onset of a research project.

However, regardless of how these lab stimuli compare to *one another*, it is also essential to assess which of them (if any) are more "realistic", or can engage with participants in a way that is similar to their real-life emotional experiences. In order to explore this, participants' subjective experiences were collected using a mobile experience-sampling application developed for Android.

Experience sampling is rooted in the vision of Brunswik (1955) - discussed in Section 1.4 (p. 49) and Section 1.4.2 (p. 52). Brunswik was an advocate of studying psychological phenomena within their natural environment, and an adversary of experimental and / or factorial designs carried out on subjects in labs. This is because, according to him, the latter method risked distorting the true relationships between psychological processes and external variables, which evolved to operate together optimally only within their natural settings. Instead, Brunswik proposed that a (now, hypothetical) researcher follow and observe the behaviour of a participant, and measure the construct of interest at random intervals, just as it occurred naturally. This idea was picked up again later by Csikszentmihalyi, Larson, and Prescott (1977), who proposed the term of "experience sampling method" for this - however this time, the venture was supported by technological advances (e.g., pagers) which allowed for the random measurement to occur without the researcher encroaching on participants' daily lives.

Since then, experience sampling has gained in popularity, and now constitutes a very welcome addition to a researcher's toolkit given its various strengths: higher statistical power due to numerous repeated measures (Hofmann & Patel, 2015); excellent compatibility with modern life given mobile phones and tablets are quasi-ubiquitous; eschewing recall and reconstruction of past experiences which is known to introduce error into measures compared to immediate report (Hofmann & Patel, 2015); and finally, providing a yardstick for other findings derived in lab settings, by either giving them more weight if they replicate in a natural environment, or to the contrary, by challenging them if they do not (Hofmann & Patel, 2015).

However, experience sampling is not without weaknesses. For instance, it may be more prone to self-selection bias compared to other research methods (Götz, Bieg, & Hall, 2016; Larson & Csikszentmihalyi, 2014); data quality may decrease over time with the repetitiveness of the task; studies may be affected by high drop-out rates due to this being a time-intensive method; and finally, this method may make participants hyper-aware of the phenomenon studied, which could affect measurement (Götz et al., 2016). But despite any drawbacks, experience sampling has proven its worth especially in contexts where a phenomenon is difficult, if not impossible, to accurately capture in a lab, e.g., the effects of solitude vs. the company of friends (Larson & Csikszentmihalyi, 2014). Moreover, it can complement lab-based results, and vice versa.

In our case, even if emotions have been studied for decades, this has most often occurred within controlled environments, whereas attempting to recreate very naturalistic testing conditions remains fairly rare. In contrast, in the current study, we designed an experience sampling phone app to easily integrate within participants' lives, in an attempt to capture emotional events as they naturally occur. The app randomly cued participants to rate any emotional event which was unfolding at the time (or within the last 30 minutes). Ratings were collected based on the same PAD measurement framework as in the previous studies in this thesis - when emotional words, sounds, images, films and VEs were used. Using this (common) framework will not only allow comparisons between real-life events and these stimulus groups, but is independently considered to be more appropriate for measuring ongoing self-reported emotional states:

> *"Self-reports of emotion are likely to be more valid to the extent that they relate to currently experienced emotions. [...] our review of this literature [...suggests] that dimensional frameworks, relative to discrete ones, better capture this measure of emotion."* (Mauss & Robinson, 2009, p. 102)

8.1.1 Aims

We aim to validate the previous elicitation methods (thus far, only compared between each other) against an external criterion: real-life emotional experiences. We also aim to assess which lab elicitation method can approximate real-life experiences most closely, if any. Furthermore, we will explore whether individual differences affect participant emotional reactions in real life, as well as the clustering structure of these reactions.

8.2 Method

8.2.1 Participants

The overall sample included 61 participants recruited mainly via the University's Careers Service, 68.85% of whom were female, with an average age across the sample of 21.66 years. The large majority of participants originated from Western Europe (45.90%), followed by the Far East (31.15%), Eastern Europe (16.39%), North America (3.27%), and Africa (3.27%).

Participants were remunerated via PayPal or with Amazon vouchers worth £15 at the end of the study, in exchange for their time commitment. As for all previous studies discussed in this thesis, participants were made aware of the exclusion criteria for this study (see Section 4.2.1), and that they required an Android phone in order to participate. Further details about the sample are provided in Table 8.1 below.

Table 8.1: Sample description based on nationality, gender, and smoking status, $N = 61$.

		Gender	
		Male	Female
Nationality	Western Europe	11	17
	Eastern Europe	2	8
	Far East	6	13
	Africa	0	2
	North America	0	2
Smoker	No	17	37
	Yes	2	5

8.2.2 Procedure

Two methods were used to recruit participants[1], following a university-wide Careers Service advertisement: participants either booked an appointment with the researcher via a SimplyBook system[2] and appeared in person at the researcher's lab, or participated fully online, never meeting the researcher.

Of the total sample of participants, 42.62% signed up for the study via the first method, and joined the researcher on a date and time of their choice, in order to fill in some questionnaire measures in OpenSesame[3], and receive both verbal and written explanations (i.e., a printed instruction pack) concerning the Android phone app they would be using over the following two weeks. Specifically, they were able to download the app and do a test-run under supervision, while following the printed app instructions in parallel (see Appendix G.1, p. 528). These were also handed over to participants, in case subsequent perusal was required.

Due to the relative inefficiency of this recruitment method and difficulties with recruitment, the researcher shifted to fully online participation for the remaining 57.38% of participants. They were able to sign up for the study online, but also answer the same questionnaire measures online in Qualtrics[4], instead of in person, in OpenSesame. In this case, participants also received standard instructions for the phone app from the researcher via a YouTube screencast of a virtual Android phone[5]. This virtual phone was emulated on a computer screen using Genymotion software[6], with the researcher demonstrating in the video how to download, install and use the phone app. Other attempts were also made to reproduce the procedure from the lab as closely as possible:

[1] The recruitment method has been coded for each participant within the collected data, and will appear in subsequent analyses as `Source` / recruitment method.

[2] Located at: https://simplybook.me/. This is an online platform allowing users or businesses to manage service bookings and avoid clashes.

[3] Downloadable from: http://osdoc.cogsci.nl/

[4] Online platform for collecting participant data. Located at: https://www.qualtrics.com

[5] https://www.youtube.com/watch?v=ZhQVgkD5zLk

[6] Downloadable from: https://www.genymotion.com/

the YouTube app demo was presented in such a way as to verbally convey the same information from the instruction pack (see Appendix G.1), and the instruction manual itself was also available as a `.pdf` download within the Qualtrics questionnaire.

Regardless of the method of recruitment, when the app was initialised for the first time post-installation, internally it would generate a notification schedule where the first day of data collection was set to the following day. The app would then run in the background on participants' phones, and notify them to rate their current emotional state 4 times randomly throughout each day, over 2 weeks, between 10am-10pm, and with a minimum gap of 2 hours between notifications. The final testing day would be the 14^{th} day after installation, when participants were instructed they could uninstall the app. Participants were also instructed to get in touch with the researcher at this time, in order to receive payment for their complete responses.

In terms of the responses required by the app within each rating session, participants were asked to: give a general description of any emotional event occurring at the time of the notification or at most 30 minutes prior to this, provide PAD ratings for the event, classify it, and specify whether they happened to be alone or in the company of others. Participants were informed that they were not required to give very detailed verbal accounts of the emotional events, and also that they can quit the study at any time if they so desire.

8.2.3 Instruments and measures

Before starting to use the app, participants completed a series of single-measure questionnaires. These included all the measures used in previous studies within this thesis (see Section 4.2.3, p. 157 for a full list), as well as several new measures.

The new measures mainly refer to whether or not participants anticipated having a couple of "eventful" weeks ahead of them (an expectation which may bias affective ratings), and how many units of alcohol they routinely consume every week (as this too could alter emotional ratings). In order to minimise the chances that participants over-interpreted their responses, the latter item was embedded within a suite of items related to general health issues, e.g., smoking and dietary habits, and exercising frequency. These newly introduced measures are printed below, alongside the variable names under which they will be known in subsequent analyses.

```
Do you expect the next 2 weeks to be typical weeks for you (i.e., no major
life events expected, like: travelling to a different country, changing
jobs, a wedding etc.)? Please provide a rating for this:
(Variable name: WeeksAhead_Gauge)
☐ 1. Only very ordinary activities planned
```

☐ 2.

☐ 3.

☐ 4.

☐ 5.

☐ 6.

☐ 7.

☐ 8.

☐ 9.

☐ 10. Life-changing events planned

How often do you exercise? (SportsFreq)

☐ Never

☐ Rarely

☐ Once a month or more

☐ Once a week or more

☐ Daily

Do you have regular meals and a healthy diet? (Diet_Meals)

☐ Often skip meals + have unhealthy snacks

☐

☐ Neither healthy, nor unhealthy diet

☐

☐ Very healthy diet + regular meal times

Do you smoke? (Smoker)

☐ Yes

☐ No

Roughly how many units of alcohol do you consume per week? (AlcUnits)

[ ]

A helping image was also included in either OpenSesame or Qualtrics to support more accurate responding to the final question above. This image included information regarding the typical number of alcohol units contained by various popular alcoholic beverages (see Figure 8.1).

Figure 8.1: Alcohol units explanatory image.

However, at the level of each phone app session, participants responded to a different set of questions. These are included in Appendix G.1, where they are reproduced as phone screenshots, in order to illustrate how these will have appeared to participants. These same questions are also listed below in summary form, in Table 8.2.

Table 8.2: Summary of items presented to participants in phone app. After filling in their user name only once during the app installation phase, the questions below would appear each time a participant was notified to submit their ratings.

Question type	Measure	Description	Low (1)	Mid (5)	High (9)
Likert, non-verbal	Valence	As before, this was implemented as a 9-point, non-verbal Likert scale. Several of the response options included additional, explanatory verbal labels:	'Extremely unpleasant'	'Neither unpleasant, nor pleasant'	'Extremely pleasant'
Likert, non-verbal	Arousal	Similarly to the Valence scale, the labels used here were:	'Extremely relaxed / bored / sleepy'	'Neutral'	'Extremely alert / agitated'
Likert, non-verbal	Dominance	See above.	'Extremely overwhelmed by the situation'	'Neither overwhelmed, nor in control'	'Extremely in control of the situation'
AffectButton	NA	After answering the SAM scales, upon waiting for approximately 1s, participants would see the AffectButton face loaded into an app screen. The AffectButton changed expression as participants dragged their finger across their phone screen, and was implemented as an `iframe` loaded directly from the researcher's GitHub repository.	NA	NA	NA
Event description	NA	Participants were also asked to provide a verbal description of event that was unfolding / had unfolded within the last 30 minutes before the app notification appeared. For this, minimal text field validation was set in place, including checking for empty responses.	NA	NA	NA
Social context	Binary	If other people were involved in the event (yes / no)	NA	NA	NA

Social context	Continuous	By implementing question logic, if a participant answered yes above, (s)he was presented with an additional question asking how many other participants were involved. Options included: 1 to 7, and over 7 people - with the latter option being coded with a value of 8.	NA	NA	NA
Event classification	NA	During each session, participants were able to create their own categories of emotional events and label them as they wished, e.g., ‘happy’, ’sad’, ’tired’. Creating one such category in a response session would ascribe the current event and its PAD ratings to this category. The next time the participant used the app, any previous categories would be displayed as reusable options, but again with the possibility of creating a new category if the previous ones were deemed unsuitable. Event category labels were restricted to be between 3 and 30 characters.	NA	NA	NA

8.2.4 Apparatus

The app's notification schedule was generated in accordance with various rules, i.e., the times for the 4 daily notifications were restricted between 10am and 10pm, and were spaced at a minimum lag of 2 hours. To achieve this, internally random noise was added to the 2 hour duration separating the notification times (but without the possibility of exceeding 4 hours), so that the specific times of individual notifications were essentially random. Across the testing period of 2 weeks, participants could snooze notifications when their timing was unsuitable - this would delay the rating procedure for 30 minutes. This option is illustrated in Appendix G.1.

Internally, the app saved the data typed in by participants into a local cache, and then sent this data onward to an online spreadsheet set up by the researcher. This cache was essential for several reasons: for the phone to keep track of the notification schedule and "know" when to prompt participants for responses, to reliably connect responses from each rating session to a given participant ID (which the participant will have entered only once during the initial app setup), and to keep track of all the previous emotional categories which the participant previously saved, in order to present them again as options.

The phone app was constructed with help from Mr. David Farrell, and by extensively altering a template offered by Thai and Page-Gould (2014, 2017)[7]. The code which supports this custom version of the app is available from the author's GitHub repository[8], which also stores the downloadable app itself[9]. Of all the necessary files involved in the correct functioning of the app, one particular JavaScript file is essential for governing its behaviour[10]. Due to fees, and difficulties with sharing and installing custom iOS apps, the app for this study was developed for Android users only.

8.3 Results

Results were computed using `R` (R Core Team, 2015), with session information (i.e., `R` version and package versions) listed in Appendix G.2, p. 541.

[7] http://www.experiencesampler.com/
[8] https://github.com/CaterinaC/Android_App
[9] https://github.com/CaterinaC/Android_App/blob/master/android-debug.apk
[10] https://github.com/CaterinaC/Android_App/blob/master/www/js/index.js.

8.3.1 Detailed sample description

Based on the recent description of measures used, the sample used in this study is characterised by the following covariate distributions - see Table 8.3 below.

Table 8.3: Android app sample description, $N = 61$.

No	Measure	Mean	Trim	Median	SD	Min	Max	Range	Skew	Kurt	SE
1	Age	21.66	21.00	20	4.04	17	36	19.00	1.48	2.03	0.52
2	Diet_Meals	2.41	2.47	2	1.13	0	4	4.00	-0.29	-0.58	0.14
3	AlcUnits	4.93	3.63	2	6.51	0	30	30.00	2.14	4.94	0.83
4	SportsFreq	2.54	2.59	3	1.01	0	4	4.00	-0.73	-0.26	0.13
5	CompComf	3.75	3.94	4	0.72	0	4	4.00	-3.52	13.12	0.09
6	CompFreq	3.97	4.00	4	0.18	3	4	1.00	-5.12	24.61	0.02
7	FilmFreq	2.48	2.49	3	0.87	1	4	3.00	-0.30	-0.76	0.11
8	GameFreq	1.72	1.65	1	1.36	0	4	4.00	0.31	-1.22	0.17
9	MMORPG Freq	0.67	0.47	0	1.00	0	4	4.00	1.47	1.40	0.13
10	PhotoFreq	2.79	2.86	3	0.97	1	4	3.00	-0.66	-0.53	0.12
11	SLFreq	0.18	0.06	0	0.50	0	3	3.00	3.47	14.50	0.06
12	VWFreq	0.75	0.49	0	1.14	0	4	4.00	1.83	2.66	0.15
13	WeeksAhead Gauge	2.26	2.00	2	2.16	0	7	7.00	0.76	-0.60	0.28
14	PANAS_Neg	5.39	5.27	5	2.71	0	12	12.00	0.33	-0.41	0.35
15	PANAS_Pos	11.18	11.29	11	3.63	4	17	13.00	-0.19	-1.04	0.46
16	PHQ_total	6.43	6.06	5	4.06	0	19	19.00	0.93	0.84	0.52
17	TAS20_DDF	8.38	8.31	8	2.79	4	14	10.00	0.26	-1.06	0.36
18	TAS20_DIF	8.07	7.53	7	6.02	0	22	22.00	0.61	-0.49	0.77

Note. **AlcUnits** = number of alcohol units consumed per week; **CompComf** = level of comfort when using a computer; **CompFreq** = frequency of using a computer; **Diet_Meals** = measure for how regular and healthy participants' meals are; **FilmFreq** = frequency of watching films; **GameFreq** = frequency of gaming; **MMORPGFreq** = Massively Multiplayer Online Role-Playing Games gaming frequency; **PANAS_Neg** = PANAS Negative Affect scale; **PANAS_Pos** = PANAS Positive Affect scale; **PhotoFreq** = frequency of taking photos; **PHQ_total** = PHQ-8 total score; **SLFreq** = frequency of using Second Life; **SportsFreq** = sports/exercise frequency; **TAS20_DIF** = TAS-20 Difficulty Describing Feelings subscale; **TAS20_DIF** = TAS-20 Difficulty Identifying Feeling subscale; **VWFreq** = frequency of using virtual worlds; **WeeksAhead_Gauge** = measure for how typical the two-week testing period is expected to be for participants' usual pattern of daily living.

Because some participants started the study in the researcher's lab, while others were recruited purely online without meeting the researcher, it was important to test for any differences due to this factor. Results for the relevant t-tests are presented in Table 8.4, where p values have *not* been corrected for multiple comparisons. This is because for tests which are incidental to the main purpose of a study, and only aim to identify possible confounders, it is preferable to be overly inclusive (Streiner & Norman, 2011), and test any covariates found within stricter subsequent models, e.g., mixed effects models (Gelman, Hill, & Yajima, 2012).

Table 8.4: Comparing continuous, single-measure, covariates between participants coming into the research lab to begin the study, vs. participating entirely online. The table represents t-test results when checking for differences between the two samples. The μ_1 column refers to participants starting the study in the researcher's lab, and μ_2 refers to participants who joined and completed the study entirely online.

No	Measure	μ_1	μ_2	Δ	t	p	df	CI lower	CI upper
1	Age	21.69	21.63	0.06	0.06	0.95	56.08	-2.03	2.15
2	Diet_Meals	2.42	2.40	0.02	0.08	0.94	45.52	-0.59	0.64
3	AlcUnits	3.35	6.11	-2.76	-1.81	0.08	54.16	-5.83	0.30
4	SportsFreq	2.81	2.34	0.46	1.86	0.07	57.90	-0.04	0.97
5	CompComf	3.81	3.71	0.09	0.54	0.59	55.80	-0.26	0.44
6	CompFreq	3.96	3.97	-0.01	-0.21	0.84	49.19	-0.11	0.09
7	FilmFreq	2.46	2.49	-0.02	-0.11	0.92	54.93	-0.48	0.43
8	GameFreq	1.62	1.80	-0.18	-0.52	0.60	53.09	-0.90	0.53
9	MMORPGFreq	0.54	0.77	-0.23	-0.97	0.34	56.98	-0.72	0.25
10	PhotoFreq	2.62	2.91	-0.30	-1.17	0.25	48.14	-0.81	0.22
11	SLFreq	0.12	0.23	-0.11	-0.95	0.35	54.71	-0.35	0.13
12	**VWFreq**	0.38	1.03	-0.64	-2.37	**0.02**	59.00	-1.19	-0.10
13	WeeksAhead_Gauge	2.77	1.89	0.88	1.58	0.12	50.90	-0.24	2.01
14	**PANAS_Neg**	6.73	4.40	2.33	3.51	**0.00**	45.75	1.00	3.67
15	**PANAS_Pos**	12.58	10.14	2.43	2.62	**0.01**	44.78	0.56	4.31
16	PHQ_total	6.96	6.03	0.93	0.88	0.38	52.83	-1.19	3.06
17	TAS20_DDF	8.88	8.00	0.88	1.23	0.23	53.59	-0.56	2.33
18	TAS20_DIF	9.19	7.23	1.96	1.25	0.22	51.12	-1.19	5.12

8.3.2 Checking participant compliance

In order to verify the level of attention that participants paid to the task, we computed a measure of compliance, i.e., inspected how many of the 4 daily measures were completed by each participant, and with what time lag after receiving the scheduled app notification. Table 8.5 below outlines the level of compliance across the sample, and Figure 8.2 (p. 350) shows that for the large majority of participants, the number of missing sessions did not exceed 15%.

Table 8.5: Participant compliance when filling in phone app items post-notifications.

No	ID	Compliance type	Testing days when user active	Sessions complete	Percentage missing sessions (relative to number of active days)	Average latency (min)
1	AF1990S	Full[a]	14	59	0.000	41.916
2	CC1990R	Full	14	64	0.000	34.079
3	CF1983S	Full	14	56	0.000	74.229
4	IB1982S	Full	14	56	0.000	73.354
5	IGM1996S	Full	14	58	0.000	27.019
6	IL1997N	Full	14	59	0.000	53.334
7	MP1997E	Full	14	56	0.000	18.285
8	NF1990S	Full	14	56	0.000	51.346
9	NS1989S	Full	14	56	0.000	55.368
10	SJ1991I	Full	14	59	0.000	57.768
11	CB1994I	Full	14	55	0.019	99.271
12	CC1998S	Full	14	56	0.019	28.489
13	CM1988C	Full	14	55	0.019	93.282
14	HS1997E	Full	14	59	0.019	72.896
15	JS1997L	Full	14	59	0.019	90.010
16	JT1994M	Full	14	56	0.019	30.764
17	MB1995E	Full	14	51	0.019	20.215
18	MY1994M	Full	14	55	0.019	13.221
19	OM1994E	Full	14	55	0.019	19.640
20	SJF1987S	Full	14	57	0.019	61.489
21	CK1994M	Full	14	56	0.037	19.662
22	DF1981S	Full	14	54	0.037	45.755
23	LCB1989S	Full	14	54	0.037	54.270
24	JO1995H	Full	14	53	0.056	42.016
25	RS1997H	Full	14	70	0.056	51.592
26	SY1998I	Full	14	54	0.056	23.268
27	HYT1994MY	Full	14	54	0.074	45.145
28	LT1995C	Full	14	51	0.093	65.275
29	MW1993S	Full	14	51	0.093	89.719
30	XN1997S	Full	14	53	0.093	80.422
31	CK1994I	Full	14	51	0.111	91.056
32	JM1996L	Full	14	50	0.111	76.436
33	DZ1996M	Full	14	61	0.130	105.132
34	LVE1997E	Full	14	48	0.148	72.459
35	SP1996T	Full	14	48	0.148	32.108
36	ME1989E	Partial[b]	12	23	0.315	56.281
37	IG1997L	Partial	11	38	0.130	75.437
38	OG1994W	Partial	10	40	0.056	56.647

39	CNV1996S	Partial	10	35	0.111	19.106
40	AW1995S	Partial	9	4	0.222	61.433
41	IR1998P	Partial	8	21	0.222	47.910
42	JL1995M	Partial	7	25	0.074	74.092
43	YT1994C	Partial	5	18	0.074	93.552
44	AC1997L	Partial	4	14	0.037	92.021
45	KM1999U	Partial	4	13	0.056	66.649
46	CM99S	Partial	3	6	0.111	17.797
47	NM1996I	Partial	2	7	0.019	16.771
48	ES1997S	Partial	2	4	0.074	47.258
49	SW1997T	Partial	1	2	0.037	107.592
50	DA1998E	None[c]	14	48	0.167	57.392
51	OJ1998I	None	14	53	0.241	68.774
52	NMK1996M	None	14	19	0.315	41.080
53	KR1995E	None	14	14	0.481	86.050
54	RN1994U	None	13	29	0.426	50.471

[a] Participant submitted ratings across all 14 days, with less than 15% of sessions missing.

[b] Participant was not active over all 14 test days, regardless of level of missingness.

[c] Participant skipped considerably more than 15% of rating sessions over the study period, and even if active for fewer than the full 14 days, could not reasonably catch up with due to severe missingness.

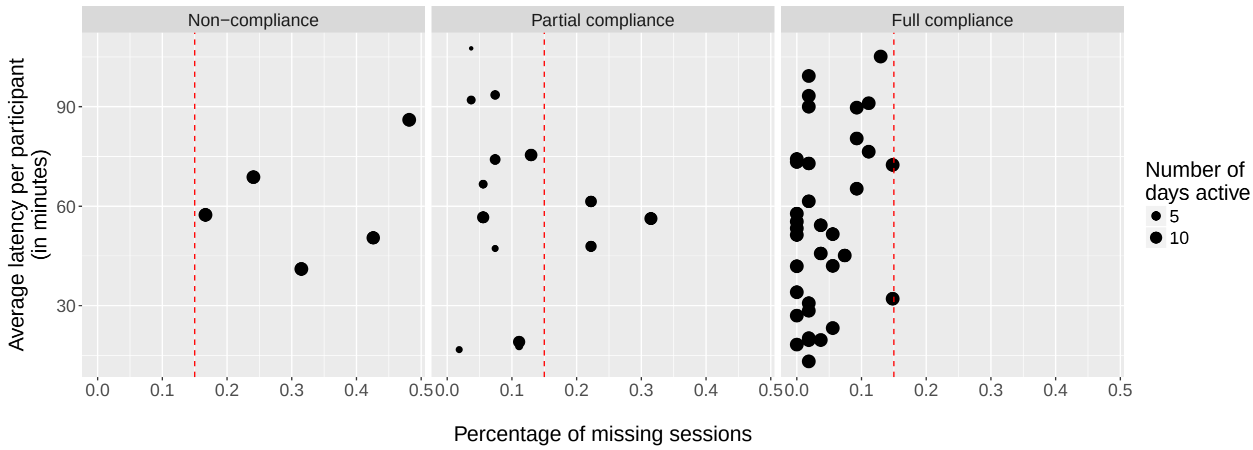


Figure 8.2: Participant compliance: percentage of skipped sessions against average time latencies.

8.3.3 Covariates affecting participant PAD ratings

In the previous Table 8.4, we have shown that three measures were influenced by the participant recruitment method (i.e., online only with questionnaires filled in using *Qualtrics* and YouTube induction, vs. face to face induction in the researcher's lab, with questionnaire responses collected in *OpenSesame*): PANAS - both Positive and Negative scores, and the frequency of using virtual worlds (VWFreq).

In this section, we further investigated whether the interaction of these variables (recruitment method × covariate measures) had any impact on participant PAD ratings. In other words, even if some covariates are significantly different between the recruitment groups, this will not necessarily translate into any differences on the outcome measures (PAD scores). Hence, we created interaction terms to test for this. If the interaction terms are non-significant, then the differences between recruitment groups in terms of the covariates are without consequence for the overall study.

For this purpose we created the mixed models below (see Table 8.6), which also include the several *newly-introduced* participant-level covariates as main effects, alongside the interaction terms of interest: e.g., weekly number of alcohol units, the extent to which participants were expecting to have an 'unusual' couple of weeks, smoking preferences, dietary habits and physical exercise frequency. One further covariate was introduced, i.e., the number of social participants involved in a given emotional event, however this was measured at the level of each phone app session, rather than at participant level. In this instance, none of the other measures logged in this study were considered, but are subsequently used in models comparing all stimulus types used in this thesis, in Section 8.3.4, which will follow.

Table 8.6: Models including newly-introduced participant covariates, which also test for interactions based on participant recruitment method. Random intercepts were added per participant within the `lmer()` R function, and all continuous predictors were standardised before being introduced into models. Interestingly, the single predictor to reach significance across all three outcomes is the number of people involved in the event which participants were rating: the larger the social group, the more positive, arousing and dominant the ratings for the event were.

Coefficients	Valence (SAM)	Arousal (SAM)	Dominance (SAM)
(Intercept)	5.81^{***}	4.94^{***}	5.33^{***}
	(0.17)	(0.12)	(0.18)
Recruitment method: Online	0.35	−0.29	0.45
	(0.25)	(0.18)	(0.27)
VWFreq	−0.27	-0.25^{*}	−0.21
	(0.15)	(0.11)	(0.16)
PANAS: Negative scale	0.03	0.09	0.01
	(0.16)	(0.11)	(0.17)
PANAS: Positive scale	0.13	0.10	0.20
	(0.15)	(0.11)	(0.16)
Smoker: Yes	−0.25	−0.54	−0.33
	(0.39)	(0.29)	(0.42)
AlcUnits	0.17	0.04	0.18
	(0.10)	(0.08)	(0.11)
Weeks ahead gauge	−0.10	−0.05	−0.14

	(0.10)	(0.07)	(0.11)
Diet & meals	0.07	0.06	0.01
	(0.10)	(0.08)	(0.11)
SportsFreq	−0.11	−0.15	−0.02
	(0.11)	(0.08)	(0.12)
Number of participants	0.16^{***}	0.09^{**}	0.11^{***}
	(0.03)	(0.03)	(0.03)
Recruitment method: Online × VWFreq	0.22	0.31^{*}	0.04
	(0.20)	(0.14)	(0.22)
Recruitment method: Online × PANAS: Negative scale	0.08	−0.28	0.08
	(0.24)	(0.17)	(0.26)
Recruitment method: Online × PANAS: Positive scale	0.05	-0.36^{*}	−0.27
	(0.21)	(0.16)	(0.23)
AIC	8821.19	8854.18	8677.94
BIC	8913.62	8946.60	8770.37
Log Likelihood	-4394.60	-4411.09	-4322.97
Num. obs.	2384	2384	2384
Num. groups: Participant	54	54	54
Var: Participant (Intercept)	0.37	0.16	0.44
Var: Residual	2.22	2.27	2.08

8.3.4 Comparing PAD ratings between 'real life' situations and all lab stimuli used

As done in previous chapters, a stepwise, backward-fitting modelling strategy was used to predict influences on PAD responses, simultaneously contrasting phone app ratings provided based on real-life situations, with ratings provided in response to lab-based stimuli. The full[11] initial models (one for each PAD dimension) included random intercepts for both individual stimuli / phone app sessions[12], as well as participants, and continuous predictors were scaled. In addition, the full models included fixed effects for the stimulus / event type (real world, words, sounds, images, film clips or VEs), as well as interactions between these stimulus types and all covariates which were common in between studies (i.e., with the exception of alcohol consumption, smoking and dietary preferences and habits, expectations concerning the weeks ahead, and sports frequency

[11] All predictors posing any interest were introduced together in a "full" model which iteratively removed terms based on maximising model fit in terms of AIC.

[12] Every time a participant was notified by the phone app that (s)he should provide ratings, this constituted a phone app *session*, which was completed in response to any emotional event occurring at the time. Such "sessions" or emotional events are analogue to individual lab stimuli for the purposes of this analysis.

- all of which were measured only for the phone app study, but were shown to not have a significant influence on ratings anyway - see Table 8.6).

For the Valence and Dominance dimensions, the baseline category referred to the ratings for real-world emotional events, against which all lab stimuli were compared. However, in the case of the Arousal dimension only, we subdivided the baseline category according to the participant recruitment method, as this was previously shown to affect ratings. In this manner, real-world ratings *from the face-to-face induction* were taken as a baseline simultaneously for all *lab stimuli*, as well as the ratings from phone app users who underwent the *online induction*. Results are presented below in Table 8.7 separately for ratings collected using the SAM scales, and the AffectButton[13]:

[13]The AffectButton was used again for measurement, however it confirmed previously found patterns of mis-alignment with the SAM (see Section 4.3.3 and Section 4.4.4). Reasons for this remain unclear, but some possibilities were advanced in Section 4.4.4. Due to this, the AffectButton results below were not discussed further.

Table 8.7: Real-life PAD ratings vs lab stimuli: SAM and AffectButton measurements.

Coefficients	Valence (SAM)	Valence (AffectButton)	Arousal (SAM)	Arousal (AffectButton)	Dominance (SAM)	Dominance (AffectButton)
(Intercept)	6.07 (0.08)***	0.25 (0.03)***	4.95 (0.17)***	−0.05 (0.10)	5.72 (0.14)***	−0.07 (0.04)
Type: film	−0.77 (0.16)***	−0.08 (0.05)	0.38 (0.22)	0.12 (0.13)	−0.89 (0.22)***	0.14 (0.06)*
Type: image	−1.26 (0.24)***	−0.18 (0.06)**	0.63 (0.26)*	0.01 (0.13)	−0.97 (0.25)***	0.10 (0.07)
Type: sound	−0.91 (0.24)***	−0.14 (0.06)*	0.31 (0.26)	−0.02 (0.13)	−0.74 (0.25)**	0.19 (0.07)**
Type: VE	−0.74 (0.18)***	−0.09 (0.05)	−0.10 (0.24)	−0.04 (0.13)	−0.44 (0.22)	0.16 (0.06)*
Type: word	−0.99 (0.24)***	−0.15 (0.06)*	0.31 (0.26)	−0.16 (0.13)	−0.69 (0.25)**	0.14 (0.07)
PANAS Positive Scale	0.10 (0.03)**					
Type: real world × Age	0.26 (0.09)**	0.05 (0.03)				
Type: film × Age	−0.07 (0.06)	−0.00 (0.02)				
Type: image × Age	−0.08 (0.06)	0.02 (0.02)				
Type: sound × Age	−0.10 (0.06)	−0.03 (0.02)				
Type: VE × Age	−0.17 (0.08)*	−0.03 (0.03)				
Type: word × Age	−0.00 (0.06)	0.01 (0.02)				
Type: real world × CompFreq		−0.00 (0.02)				
Type: film × CompFreq		−0.02 (0.02)				
Type: image × CompFreq		−0.05 (0.02)*				
Type: sound × CompFreq		−0.04 (0.02)				
Type: word × CompFreq		−0.00 (0.02)				
Type: real world (Face to face induction)			−0.13 (0.24)	0.02 (0.14)		
Type: real world (Face to face induction) × GameFreq			−0.01 (0.18)			
Type: film × GameFreq			−0.01 (0.12)			
Type: image × GameFreq			0.05 (0.13)			
Type: real world (Face to face induction) × GameFreq			0.01 (0.17)			
Type: sound × GameFreq			−0.22 (0.13)			
Type: VE × GameFreq			−0.04 (0.13)			
Type: word × GameFreq			−0.21 (0.13)			
Type: real world (Face to face induction) × MMORPGFreq			−0.26 (0.25)			
Type: film × MMORPGFreq			−0.25 (0.15)			−0.01 (0.05)

Table 8.7: Real-life PAD ratings vs lab stimuli: SAM and AffectButton measurements (continued).

Coefficients	Valence (SAM)	Valence (AffectButton)	Arousal (SAM)	Arousal (AffectButton)	Dominance (SAM)	Dominance (AffectButton)
Type: image × MMORPGFreq			−0.26 (0.16)			0.04 (0.05)
Type: real world (Face to face induction) × MMORPGFreq			0.00 (0.13)			
Type: sound × MMORPGFreq			−0.10 (0.16)			−0.04 (0.05)
Type: VE × MMORPGFreq			0.14 (0.10)			0.03 (0.03)
Type: word × MMORPGFreq			−0.09 (0.16)			−0.02 (0.05)
Type: real world (Face to face induction) × SLFreq			0.04 (0.19)			
Type: film × SLFreq			0.03 (0.11)			
Type: image × SLFreq			0.18 (0.11)			
Type: real world (Face to face induction) × SLFreq			0.07 (0.19)			
Type: sound × SLFreq			0.03 (0.11)			
Type: VE × SLFreq			−0.11 (0.12)			
Type: word × SLFreq			0.21 (0.11)			
Type: real world (Face to face induction) × Nationality: Far East				0.13 (0.16)		
Type: film × Nationality: Far East				0.07 (0.12)	−0.14 (0.24)	
Type: image × Nationality: Far East				0.14 (0.12)	−0.20 (0.26)	
Type: real world (Face to face induction) × Nationality: Far East				0.23 (0.17)		
Type: sound: Nationality: Far East				−0.03 (0.12)	−0.34 (0.26)	
Type: VE × Nationality: Far East				0.12 (0.12)	−0.12 (0.26)	
Type: word × Nationality: Far East				0.25 (0.12)*	−0.57 (0.26)*	
Type: real world (Face to face induction) × Nationality: Eastern Europe				−0.15 (0.22)		
Type: film × Nationality: Eastern Europe				−0.10 (0.11)	0.17 (0.23)	
Type: image × Nationality: Eastern Europe				0.04 (0.11)	−0.02 (0.24)	
Type: real world (Face to face induction) × Nationality: Eastern Europe				−0.29 (0.19)		
Type: sound: Nationality: Eastern Europe				0.03 (0.11)	0.06 (0.24)	
Type: VE × Nationality: Eastern Europe				−0.11 (0.13)	−0.13 (0.28)	
Type: word × Nationality: Eastern Europe				0.07 (0.11)	−0.03 (0.24)	

Table 8.7: Real-life PAD ratings vs lab stimuli: SAM and AffectButton measurements (continued).

Coefficients	Valence (SAM)	Valence (AffectButton)	Arousal (SAM)	Arousal (AffectButton)	Dominance (SAM)	Dominance (AffectButton)
Type: film × Nationality: North America				0.05 (0.16)	0.10 (0.35)	
Type: image × Nationality: North America				0.33 (0.17)	−0.55 (0.36)	
Type: real world (Face to face induction) × Nationality: North America				−0.71 (0.30)*		
Type: sound: Nationality: North America				0.17 (0.17)	0.17 (0.36)	
Type: VE × Nationality: North America				−0.06 (0.14)	−0.28 (0.30)	
Type: word × Nationality: North America				0.06 (0.17)	−0.01 (0.36)	
Type: film × Nationality: Middle East				0.00 (0.34)	0.37 (0.72)	
Type: image × Nationality: Middle East				0.05 (0.35)	−0.15 (0.76)	
Type: sound: Nationality: Middle East				0.15 (0.35)	−0.38 (0.76)	
Type: VE × Nationality: Middle East				−0.45 (0.21)*	−0.34 (0.44)	
Type: word × Nationality: Middle East				0.08 (0.35)	0.09 (0.76)	
Type: film × Nationality: South America				0.18 (0.36)	0.53 (0.72)	
Type: image × Nationality: South America				0.08 (0.37)	−0.15 (0.76)	
Type: sound: Nationality: South America				0.11 (0.37)	−0.66 (0.76)	
Type: VE × Nationality: South America				−0.43 (0.35)	0.03 (0.73)	
Type: word × Nationality: South America				0.17 (0.37)	0.93 (0.76)	
Type: real world (Face to face induction) × Nationality: Africa				0.25 (0.36)		
Type: VE × Nationality: Africa				−0.62 (0.34)	0.31 (0.73)	
Type: VE × Nationality: Australia				−0.24 (0.24)	−0.96 (0.52)	
Type: real world (Face to face induction) × PHQ_total				−0.01 (0.07)		
Type: film × PHQ_total				−0.03 (0.05)		
Type: image × PHQ_total				−0.02 (0.05)		
Type: real world (Face to face induction) × PHQ_total				0.14 (0.10)		
Type: sound × PHQ_total				0.06 (0.05)		
Type: VE × PHQ_total				−0.05 (0.05)		
Type: word × PHQ_total				0.09 (0.05)		

Table 8.7: Real-life PAD ratings vs lab stimuli: SAM and AffectButton measurements (continued).

Coefficients	Valence (SAM)	Valence (AffectButton)	Arousal (SAM)	Arousal (AffectButton)	Dominance (SAM)	Dominance (AffectButton)
Type: real world (Face to face induction) × TAS20_DDF				−0.04 (0.08)		
Type: film × TAS20_DDF				0.04 (0.04)		
Type: image × TAS20_DDF				0.05 (0.05)		
Type: real world (Face to face induction) × TAS20_DDF				−0.20 (0.09)*		
Type: sound × TAS20_DDF				0.02 (0.05)		
Type: VE × TAS20_DDF				0.02 (0.05)		
Type: word × TAS20_DDF				−0.03 (0.05)		
Type: real world × Nationality: Far East					−0.43 (0.23)	
Type: real world × Nationality: Eastern Europe					−0.33 (0.31)	
Type: real world × Nationality: North America					0.64 (0.56)	
Type: real world × Nationality: Africa					−0.42 (0.76)	
Type: real world × FilmFreq						−0.05 (0.04)
Type: film × FilmFreq						0.00 (0.04)
Type: image × FilmFreq						−0.05 (0.04)
Type: sound × FilmFreq						−0.00 (0.04)
Type: VE × FilmFreq						−0.01 (0.04)
Type: word × FilmFreq						−0.05 (0.04)
Type: real world × MMORPGFreq						−0.03 (0.03)
AIC	51268.52	18733.45	56500.87	29071.71	53009.47	21143.56
BIC	51389.71	18884.94	56735.68	29503.45	53312.45	21302.63
Log Likelihood	-25618.26	-9346.72	-28219.44	-14478.85	-26464.74	-10550.78
Num. obs.	14393	14393	14393	14393	14393	14393
Num. groups: EventOrStimulus	2585	2585	2585	2585	2585	2585
Num. groups: PID	173	173	173	173	173	173
Var: EventOrStimulus (Intercept)	1.13	0.07	0.69	0.05	0.47	0.05
Var: PID (Intercept)	0.19	0.03	0.64	0.10	0.47	0.07
Var: Residual	1.75	0.19	2.64	0.40	2.09	0.23

8.3.5 Population equivalence between real-world and lab stimuli PAD ratings

Given that the mixed models above often flagged discrepancies between the real-world PAD ratings and the other stimulus modalities, we further tested this discrepancy using the Kolmogorov-Smirnoff test. While this test is routinely used to assess distribution normality, this is only a special case of its wider applications: to test whether two samples/distributions are likely to have been drawn from the same (but not necessarily known) statistical population (see I. T. Young, 1977). When the cumulative distribution functions describing the two samples differ too widely, the test returns a significant result. Using this line of reasoning, we verified whether any of the lab stimuli produced distributions that were likely to come from the same statistical population as the real-world ratings (see Table 8.8).

However, given that the Kolmogorov-Smirnoff test is very sensitive to differences around the medians of the two distributions, and not the tails (see Engmann & Cousineau, 2011), an additional test was sought which could also fulfil this condition: the Anderson-Darling test (see results for this test in Table 8.9). We found that, regardless of the test used and the PAD dimension assessed, all results emerged as significant, suggesting that the underlying statistical populations for lab stimuli were indeed different to those of real-world PAD ratings, i.e., were described by different cumulative distribution functions.

Table 8.8: Bootstrapped Kolmogorov-Smirnoff test, with 10000 samples. In all cases, this test suggests that the distributions are part of different populations.

Dimension	Comparison	KS statistic	p
Valence	Real world vs emotional words	0.214	0
	Real world vs emotional sounds	0.170	0
	Real world vs emotional images	0.246	0
	Real world vs emotional film clips	0.172	0
	Real world vs emotional VEs	0.150	0
Arousal	Real world vs emotional words	0.139	0
	Real world vs emotional sounds	0.173	0
	Real world vs emotional images	0.234	0
	Real world vs emotional film clips	0.197	0
	Real world vs emotional VEs	0.074	0
Dominance	Real world vs emotional words	0.156	0
	Real world vs emotional sounds	0.150	0
	Real world vs emotional images	0.202	0
	Real world vs emotional film clips	0.149	0
	Real world vs emotional VEs	0.096	0

Table 8.9: Anderson-Darling test results. In all cases, these results suggest that real world PAD ratings are part of different populations, relative to ratings of emotional lab stimuli.

Dimension	Comparison	AD[a]	T.AD[b]	p	N_1[c]	N_2[d]
Valence	Real world vs. emotional words	98.78	128.48	0.00	1500	2384
	Real world vs. emotional sounds	79.02	102.52	0.00	1500	2384
	Real world vs. emotional images	131.64	171.67	0.00	1500	2384
	Real world vs. emotional films	84.00	109.04	0.00	4500	2384
	Real world vs. emotional VEs	82.71	107.35	0.00	3009	2384
Arousal	Real world vs. emotional words	54.62	70.45	0.00	1500	2384
	Real world vs. emotional sounds	81.78	106.14	0.00	1500	2384
	Real world vs. emotional images	133.76	174.45	0.00	1500	2384
	Real world vs. emotional films	168.44	219.97	0.00	4500	2384
	Real world vs. emotional VEs	39.00	49.92	0.00	3009	2384
Dominance	Real world vs. emotional words	66.05	85.47	0.00	1500	2384
	Real world vs. emotional sounds	69.01	89.37	0.00	1500	2384
	Real world vs. emotional images	114.01	148.49	0.00	1500	2384
	Real world vs. emotional films	106.16	138.15	0.00	4500	2384
	Real world vs. emotional VEs	35.20	44.94	0.00	3009	2384

[a] Anderson-Darling (AD) test statistic.
[b] Standardised AD test statistic.
[c] Sample sizes for the various types of lab stimuli.
[d] Sample size of the real-world PAD ratings gathered using the Android app.

8.3.6 Verbal labels used by participants to classify real-life emotional situations

We also investigated the verbal labels under which participants classified their daily emotional events. These could be meaningful in themselves, but more importantly, they could serve as a higher level at which to aggregate the PAD ratings (i.e., using the continuous and averaged PAD ratings across these participant categories for subsequent analyses, rather than integer PAD ratings from single app sessions). We thus proceeded to lemmatise and stem all category labels / terms, and then individually compute their frequencies. After excluding any words with unknown lemmas (e.g., 'meh'), words with a frequency of 1, and words with less than 3 characters (i.e., stop-/connector words), the remaining terms were plotted as a wordcloud in Figure 8.3 (p. 360).

8.3.7 Clustering real-life emotional events, compared to lab stimuli

We were able to compare the phone-app PAD averages to the average ratings collected across all previously tested lab stimuli (be it words, sounds, images, films or VEs). More specifically, for real life, individual-session PAD ratings were averaged at the level of the emotional category the event was placed in (with participant-defined categories having

Figure 8.3: Wordcloud based on individual terms used by participants to label their phone app emotional categories. They seem to cover a fairly wide range of content, and emotional words appear to be the most common, relative to any non-emotional words.

been discussed above, and illustrated in Figures 8.4 and 8.5, p. 362-363). Hence, the following analyses rely on the category-wide PAD averages for emotional events, pooled across all categories created by the various participants using the phone app. For the lab stimuli, each stimulus was also represented by the average PAD values collected from a full sample of participants.

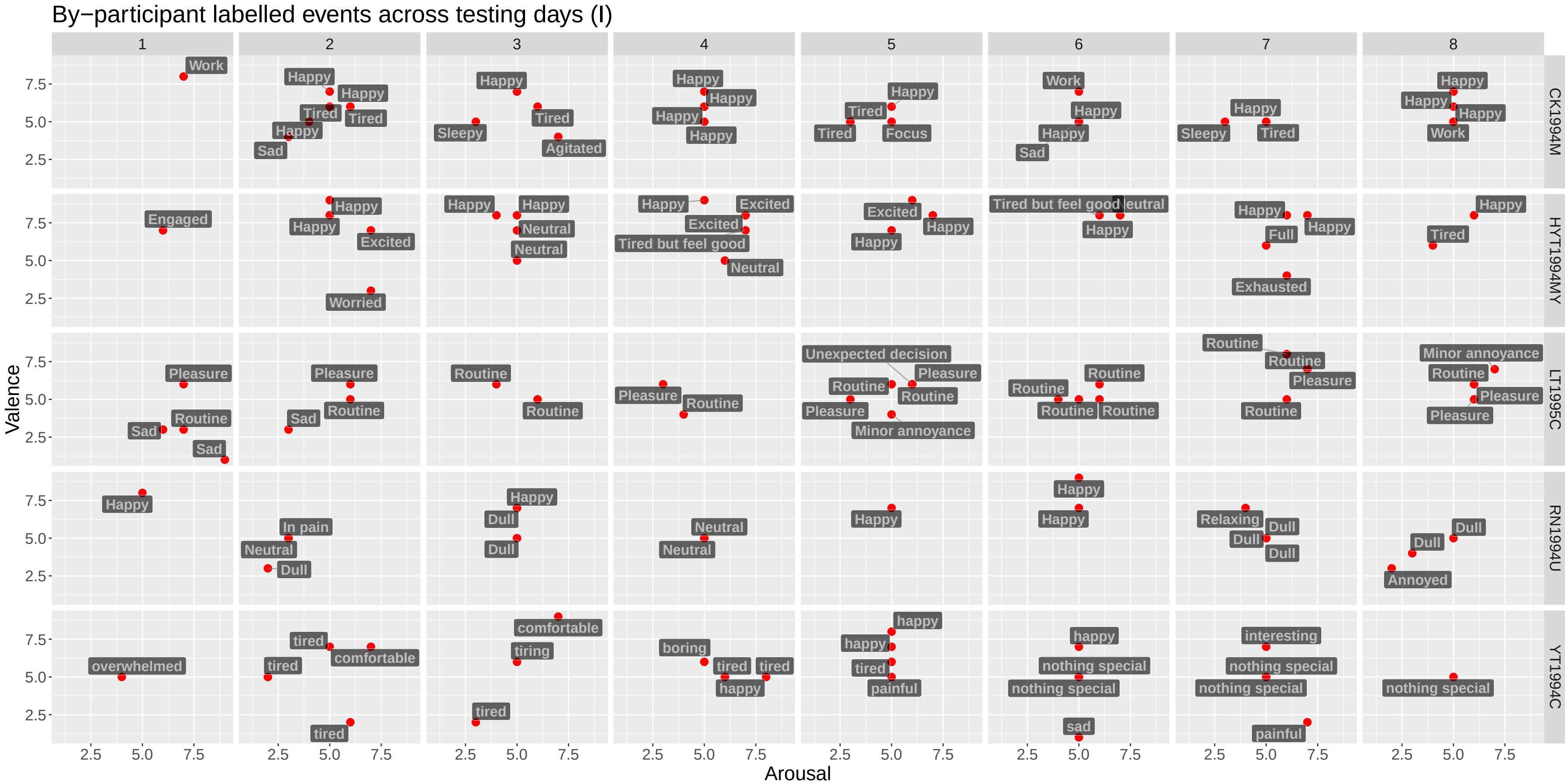


Figure 8.4: Participant categories labelled by day, for days 1 to 8. PAD ratings were averaged according to each of these category labels, e.g., all events submitted under the category of "Sad" were averaged in terms of Valence, Arousal, and Dominance, with these constituting the three coordinates of a single data point, for later using in a MBC analysis.

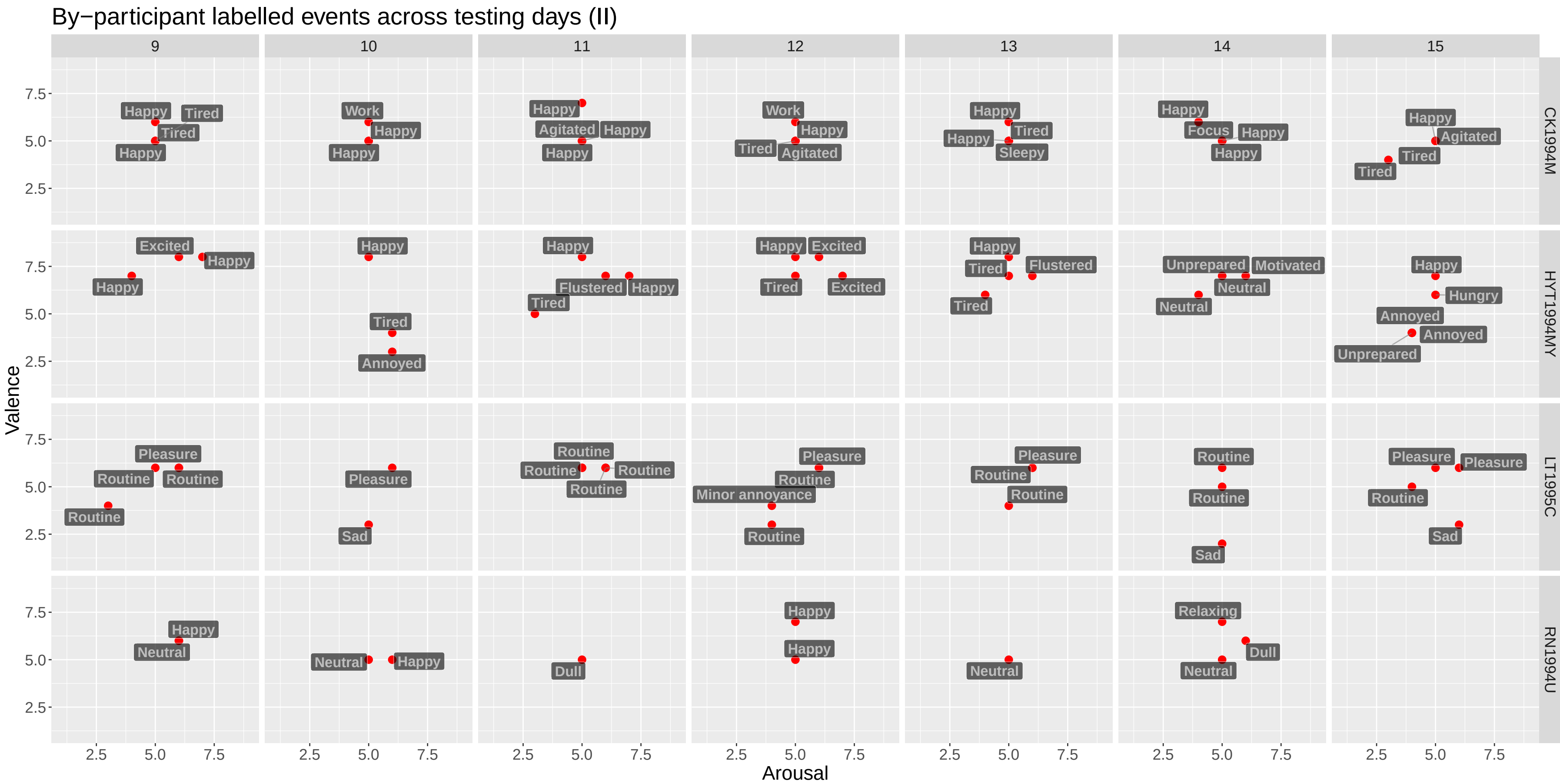


Figure 8.5: Participant categories labelled by day, for days 9 to 15. An extra day of testing is present for participants who also submitted ratings as part of their practice.

Because emotional categories from the phone app, as well as the stimuli intended for lab use, can both be considered types of emotional ‘events’, or ‘triggers’, in the following we will be comparing the average PAD ratings across these two conditions. This comparison was explored both visually (see Figure 8.6, p. 365), and via model-based clustering (MBC) below. It is worth highlighting that data pertaining to *all* types of lab-stimuli (i.e., words, sounds, images, films and VEs) was fed to the MBC algorithm, in order to map out the maximum amount of PAD space covered by all these lab stimuli simultaneously[14].

After an MBC algorithm was applied to each of these aggregated datasets, we discovered that the phone app / real-world data was best described by a model with k = 3 components, with ellipsoidal distributions and *variable* volume, and equal shape and orientation (VEE), whereas the lab data - by a model with $k = 6$ components, also with ellipsoidal distributions, but *equal* volume, shape and orientation (EEE). Therefore, not only does the number of clusters vary considerably between the two streams of data (i.e., lab-based vs. real world), but the characteristics of the clusters differ as well. Centroids for each solution are presented below in Table 8.10, with the classification for each also being illustrated in Figure 8.7 (p. 366).

Table 8.10: MBC centroids and mixing proportions for the real world categories and lab stimuli.

Source	Cluster	Valence	Arousal	Dominance	Mixing proportion
Phone app	1	7.08	5.09	6.19	0.449
	2	3.46	4.99	3.82	0.235
	3	5.31	4.95	5.08	0.315
Lab stimuli	1	6.72	6.01	5.74	0.153
	2	5.13	3.69	4.89	0.081
	3	6.38	4.25	5.59	0.311
	4	2.91	6.70	3.57	0.203
	5	6.77	7.58	4.85	0.015
	6	4.61	5.20	4.70	0.237

In terms of model fit and BIC values, the optimal model identified for the real world data (k = 3, VEE) differed only slightly from the next best fitting option by 0.505 BIC points (i.e., also $k = 3$, for an EVE model - with ellipsoidal distributions, of equal volume and orientation, but variable shape). In fact, the top three models were all k = 3 solutions, but of varying shapes. In this case, BIC values were distributed between a

[14] To read about clustering solutions developed for a single or fewer modality/-ies, please see e.g., **Section 2.3.2** for image classification results, **Section 3.3.3** for images matched to words and sounds, **Subsection 4.3.4.1** for images, sounds, words and films, and both **Section 5.3.5** and **Section 7.3.7** for VEs. Findings range from a minimum of $k = 2$ (for VEs alone), to a maximum of $k = 5$ for images alone.

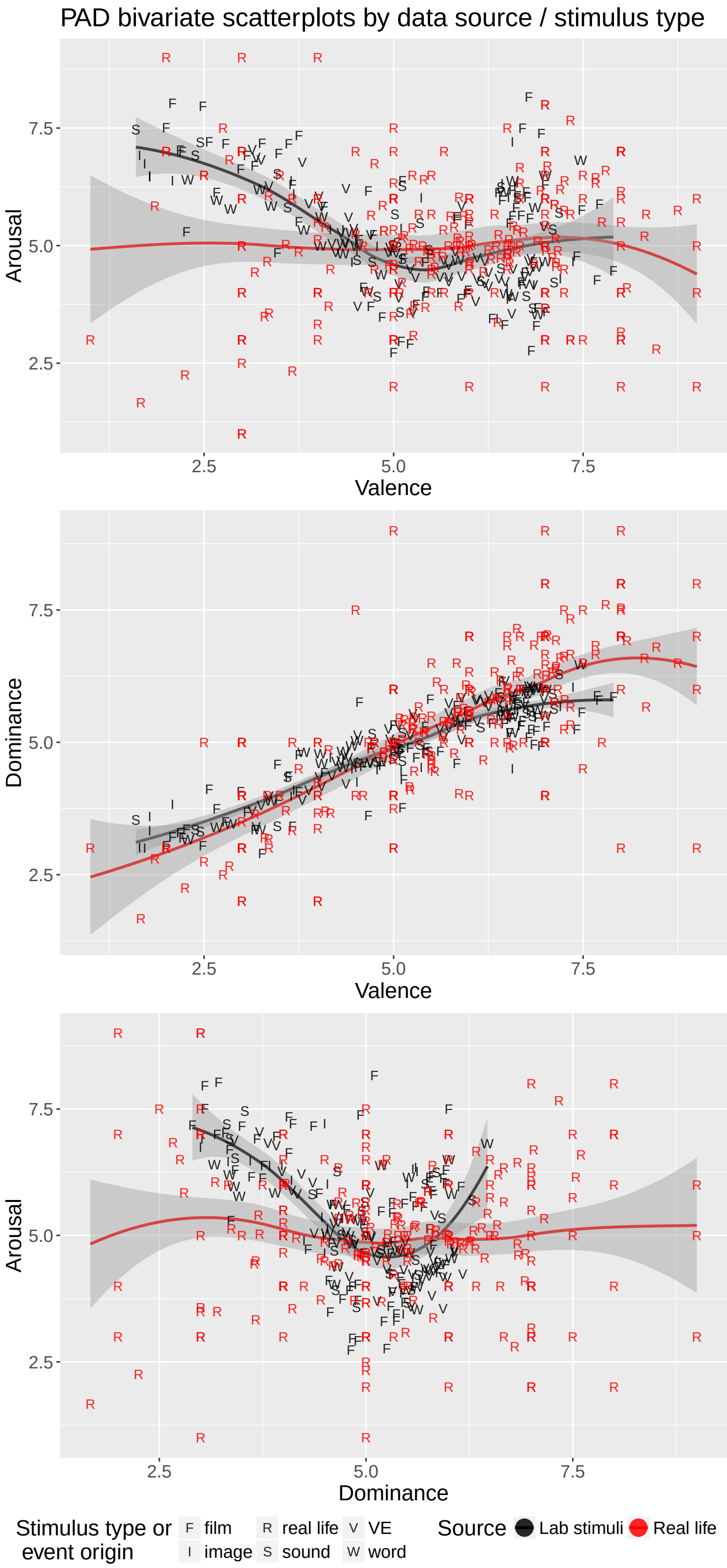
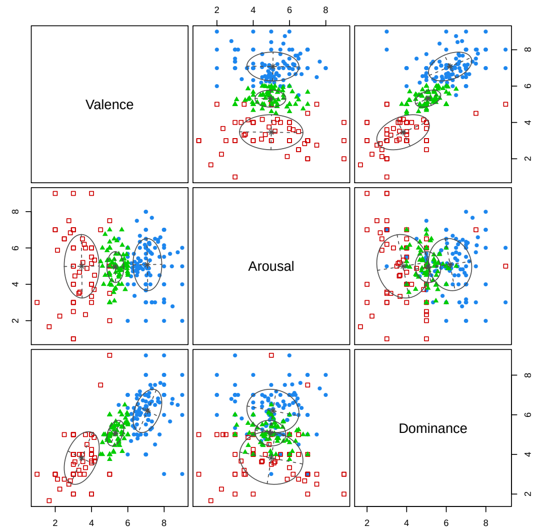
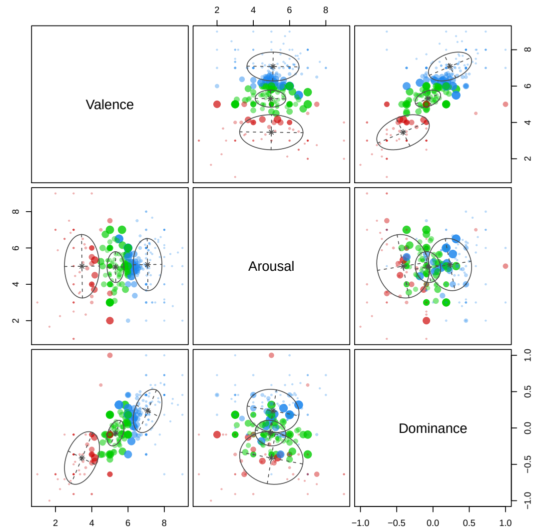


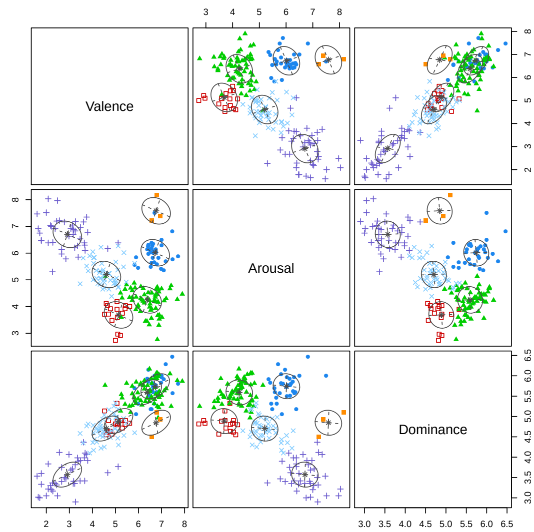
Figure 8.6: PAD scatterplots comparing real-world ratings to lab stimuli. Several observations emerge here: that the lab stimuli, regardless of modality, do not cover an area of emotional content as wide as the real-life events (i.e., *range restriction*), and that the relationships between PAD dimensions differ in *shape* depending on the source of the data (i.e., quadratic trends seen for lab stimuli are absent for real-world emotional data). This plot also lends more weight to earlier results seen from the KS and AD tests - namely that these values do not originate from the same populations.



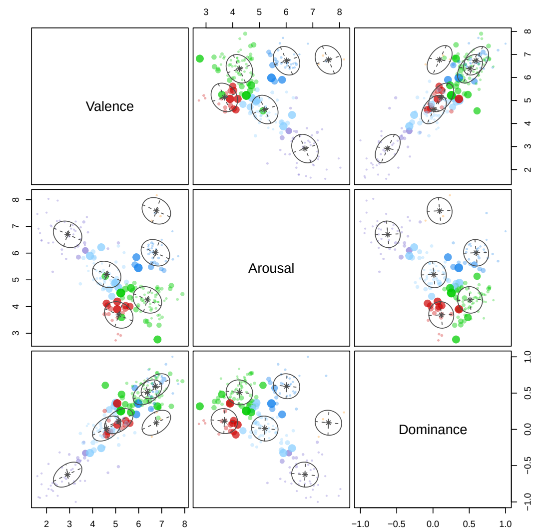
(a) Classification for **real life emotional categories** into $k = 3$ clusters.



(b) Real world events: membership uncertainty. The larger and more opaque the points, the higher the uncertainty.



(c) Classification for pooled **lab stimuli** into $k = 6$ clusters.



(d) Pooled lab stimuli: membership uncertainties for the MBC solution. The larger and more opaque the points, the higher the uncertainty.

Figure 8.7: Real world emotional events vs. lab stimuli MBC solutions: memberships and uncertainties.

maximum of -2566.53, and a minimum of -2767.00 for the poorest fitting model, with a median and mean of -2613, and -2629, respectively. In the case of the lab stimuli, the optimal classification (k = 6, EEE model) was separated from the next closest fit (k = 8, given up to 9 components being tested) by 1.43 BIC units, and 535 units from the poorest. This distribution of BIC values had a median and mean of -1480 and -1514, respectively.

8.4 Discussion

The most striking finding within our models (which pooled real-world data and data from all lab stimulus modalities), was that the Valence ratings associated with real-life emotional events unanimously emerged as more positive than any of the lab stimulus modalities tested (when measured using the SAM). This suggests that lab research may underestimate the frequency of perceived positive content in natural environments, as it tends to favour designs with equal numbers of negative as positive stimuli. In terms of Dominance, interestingly, the lab stimuli again emerged almost always to be less confidence-, and control-inducing compared to to real-life, where participants appear to feel more emotionally secure (at least based on these data). The only exception here were VEs, which did not differ from real-life in terms of Dominance.

The fewest differences between real-life and lab ratings occurred for the Arousal dimension - where only images emerged as significantly more arousing than real-life, in contrast to all other modalities, which were as arousing as real-life. In addition, across PAD dimensions, VEs were relatively-speaking the most similar lab stimulus to real-life experiences, in that they did not differ significantly from real-life on Arousal or Dominance, whereas all other modalities did so significantly, on either one or both of these PAD dimensions. However, it is perhaps risky to read too much into such null results, particularly since VEs also underestimated the frequency of positive events (as did all the other modalities).

One effect which could not be tested across modalities is the influence of social context. The affective words, sounds, images, films and VEs used in lab settings are almost always used on isolated participants, hence introducing another important departure from real-life emotional experiences. In fact, in this study, we found that emotional experiences were rated as more positive, the larger the group that participants were in, similar to the findings of Larson and Csikszentmihalyi (2014), who suggest that presence within a group can support emotion regulation and "lift participants' mood" after a negative event. Not only this, but emotional events also tended to be more arousing, and more Dominance-inducing when participants were in the company of a group. This may be also be related to effects of mood contagion (Barsade, 2002; Neumann & Strack, 2000), or perhaps impression management by individuals when they are within a group (Leary & Kowalski, 1990).

When checking for any evidence of population equivalence between the real-world and lab-based PAD ratings, the search for a lab stimulus modality capable of mimicking real-life experiences returned empty: both the Kolmogorov-Smirnoff and Anderson-Darling tests suggested that PAD ratings for real-life vs. lab stimuli are drawn from different statistical populations, irrespective of the PAD dimension or lab stimulus concerned. This introduces essential questions regarding the interpretability of lab-based results, if or when researchers do not attempt to corroborate these with real-world data.

Structural differences also set real-life and lab data apart further: real-life emotional episodes were shown to cover a far wider area of PAD space (especially in terms of Arousal and Dominance) than any of the lab stimuli assessed (including VEs). Not only this, but the bivariate shape of PAD relationships was different across real-life emotional situations and lab stimuli. The (by now) familiar quadratic trend between Valence and Arousal was only present for lab stimuli, and collapsed into a near-flat line for real-life data - indicating orthogonality between these two dimensions instead. This is a situation where using lab stimuli as proxies for real-life experiences in order to estimate the shape of the Valence $\times$ Arousal relationship could be grossly misleading. The same was true of the bivariate relationship between Dominance and Arousal: the quadratic trend present for lab stimuli again vanished for real-life emotional ratings.

It was only the relationship between Dominance and Valence that was similar regardless of the source of the ratings - with a positive, linear relationship emerging for both lab stimuli and real-life data. However even here, range restriction affected lab stimuli in an obvious manner, as they covered less PAD space compared to real experiences. Also, real-life ratings were interestingly marked by a clear case of heteroscedasticity, especially at higher levels of Valence and Dominance - suggesting that positive and reassuring / confidence-inducing, real emotional experiences are far more diverse than those captured in lab research.

Furthermore, clustering PAD ratings from both sources of data (lab stimuli vs. real-life data collected using the phone app) enforced the differences between them. The MBC solution closely followed the linear relationship between PAD dimensions for the real-world data, whereas the lab data formed clusters which neatly followed the obvious "U"-shape between Valence and Arousal. Other differences were shown to exist as well: only $k = 3$ large clusters defined the phone app data, whereas the clusters emerging from lab stimuli were both smaller and more diverse ($k = 6$).

Across the two separate clustering solutions, an almost perfectly neutral cluster emerged naturally - but only for real-life data (i.e., all PAD centroids were extremely close to a value of 5, which is midpoint of the SAM scales), whereas for lab data, the cluster closest to being "neutral" only presented medium Valence, but lower Arousal and Dominance. This discrepancy also calls back to previous findings concerning IAPS clusters (see Table 2.2, p. 81), which could similarly have led one to believe that neutrality is defined by medium Valence, but lower Arousal and / or Dominance. And yet, this may merely be the result of lab stimuli not adequately sampling real-life emotional content, and introducing error into research conclusions.

As a reminder, the method of recruiting participants had to be altered after the onset of the study for legitimate reasons. This, however, introduced a certain amount of error in results: the online-only half of the sample was shown to use online virtual worlds significantly more often than the participants who visited the researcher's lab, and who received a face-to-face induction. Also, participants invited to the lab showed an interesting pattern of significantly higher general positive *and* negative affect (on the PANAS Positive and Negative Affect scales) than the online group. These differences translated into two significant interaction effects: between the recruitment method and virtual world usage, on the one hand, and between the recruitment method and the positive PANAS scale, on the other - both with repercussions for the Arousal scale only. For instance, participants recruited and inducted online, who also used virtual worlds more frequently, tended to report higher Arousal levels over the two weeks of testing. Also within the same online group, the higher the general positive affect, the lower the Arousal ratings submitted. Virtual world usage use also had a *main* effect on Arousal, regardless of the recruitment group, with more frequent use of virtual worlds being associated with lower Arousal ratings.

Be that as it may, the impact of these relationships was limited to only the current study. When pooling this data with all five previously collected modalities of lab stimuli (i.e., affective words, sounds, images, films and virtual environments), this effect became non-significant. More specifically, we investigated this by using the various lab stimulus modalities as predictors for the PAD ratings provided by participants, with the real-world category serving as a baseline in the model. While this format was appropriate for predicting Valence and Dominance, a more detailed strategy was employed for Arousal due to the interactions flagged above. Hence, the 'real-world' data were split into two separate subcategories, depending on the method of recruiting participants (fully online vs. face-to-face induction in the lab). Then, the interactions between recruitment method and the PANAS scales / virtual world frequency were re-assessed, but this time no longer reach significance when predicting Arousal. From this, we concluded that, although changing the method of participant recruitment during the study was far from ideal, its impact on results could be ignored.

Finally, it is worth mentioning that the level of participant compliance for this task was adequate in terms of the number of sessions submitted overall. However, participants tended to show anywhere between 13.22 and 107.59 minutes of latency (on average), in between the pre-determined notification time for their rating session, and the moment they actually submitted their ratings. It is widely agreed that the sooner participants fill in the items after being cued, the more valuable the data (Csikszentmihalyi & Larson, 2014; Larson & Csikszentmihalyi, 2014). In this case, it is unclear whether participants saw the notifications but ignored them and decided to respond later (in which case, it is possible for recall to have influenced ratings to some degree), or whether they never noticed the notifications until later (e.g., if the phone was in silent mode etc.), and then immediately rated any emotional event that was occurring at the time.

8.4.1 Limitations

A few limitations concerning the sampling policies in this study may affect inference: ideally, participant nationality should have been kept constant across the sample, due to known differences in emotion processing (e.g., Jack et al., 2009). This was not possible however, given that it was particularly difficult to recruit 'adequate' numbers of participants willing to commit to a two-week long study, with payment offered at the end.

Furthermore, most of this sample was composed of female participants, which may also limit the generalisability of results. In addition, for objective reasons (including lower costs), our experience-sampling app was developed for Android users only, which implicitly excluded Apple users from the study. Given the widespread popularity of both phone operating systems, it is debatable whether this will necessarily have skewed results.

Finally, a clear limitation of this study is that the recruitment method was altered half-way through data collection, which was shown to have some impact in secondary analyses. Fortunately, for major models comparing all five lab stimulus modalities with real-world data, this was shown to not have a significant impact.

8.4.2 Future directions

A more powerful design could be created for the same purpose as the current study. However, this would require the same participant sample to take part in the experience-sampling task, as well as provide ratings for the various lab stimulus modalities. If solutions can be found to reduce the considerable time commitment for these participants, such a design would be invaluable in mapping with more confidence the areas of PAD space sampled by real-life experiences, but not lab stimuli - or perhaps vice-versa. In addition, future research could also investigate the extent to which emotional research results in the lab may be biased by the fact that they often involve isolating participants from any a social context.

Part VI

General discussion

Chapter 9

General discussion

In this chapter, key findings will first be reiterated briefly, and then any central themes extracted from this doctoral research will be discussed in the wider context of the literature. The chapter will conclude with directions for future research, and limitations.

9.1 Summary of key findings

IAPS image selection

In this work, we proposed a standard, replicable method for selecting IAPS stimuli using model-based cluster analysis (MBC). The method included screening the norms of the IAPS database for outliers, identifying a suitable clustering solution, and extracting stimuli for research on the basis of their level of uncertainty.

Key findings included that a five-cluster solution (with neutral, mildly negative, intensely negative, positive and serene, and positive and exciting clusters) was well-supported on this image dataset - a departure from previous researchers' work, which usually use only 3 categories. Of these 5 clusters, the one aligned most closely to the idea of "neutrality" presented medium Valence, but low Arousal, and high Dominance[1].

Other findings derived from this study included the "U"-shaped relationship between Valence and Arousal, alongside the limitations this imposes on the process of sampling stimuli. For instance, if a factorial design was used with these IAPS stimuli in the absence of an orthogonal relationship between these dimensions, it would be difficult to populate the cells for every combination (above and beyond the arguments forwarded by Brunswik in 1955 against using factorial designs at all). The reason why this quadratic trend occurred instead of an orthogonal relationship is not clear, and any less (or un)populated areas of PAD space could

[1] The issue of neutrality, and its importance, will be discussed in the upcoming Section 9.2.2 in more depth, and in the light of findings from subsequent chapters.

indicate that the IAPS database is either incomplete, or perhaps that *image* stimuli are not very suitable for representing all PAD combinations, compared to other modalities.

Finally, we have also argued that cluster analysis greatly facilitates the simultaneous consideration of all PAD dimensions within the stimulus selection process, compared to other "manual" methods. Hence, by jointly incorporating Valence, Arousal, and Dominance into our analysis, the principal output of this study was a list of over 800 IAPS images, classified into the 5 clusters mentioned above, and sorted by their uncertainty of belonging to their respective cluster (see Appendix A).

Matching images, words and sounds

This study involved a matching process whereby each sound was coupled with both a word and an image, before running a model-based cluster analysis on the resulting matched triplets. An imperfect degree of overlap in PAD space was found between the three stimulus modalities, resulting later in the collapse of the neutral cluster (hence, $k = 4$ here). This occurred because the IADS-2 sound database (which the entire matching process relied on, being the smallest) did not present enough medium Valence and low Arousal sounds - the combination typical of "neutral" material in the IAPS (and ANEW). To compensate for this an artificial neutral cluster was created, however this substitute was designed instead to be as close as possible to the centre of PAD space (i.e., a location defined by the midpoints of the Valence, Arousal and Dominance).

The fact that the three modalities did not show perfect overlap in terms of their PAD coordinates casts some doubt as to whether a "central emotional system" exists for dealing with all types of emotional stimuli[2]. It also suggests that choosing just one modality over others for research may skew results. In fact, it was noted that the matching process between modalities also heavily shifted the position of best representatives between the IAPS-only clustering solution, and the mixed-modality solution. This is further evidence against a parsimonious, "central emotional system". Despite this, the "U"-shaped relationship between Valence and Arousal re-emerged for all three stimulus modalities[3].

For this and the previous study, it is worth highlighting the benefits of using data-driven methods (i.e., model-based clustering) for selecting stimuli:

- Reducing the amount of subjectivity / arbitrary decisions within the stimulus selection process;
- Taking all three PAD dimensions into account simultaneously, and in an effortless manner compared to manual methods;

[2] This theme will be developed more shortly, in Section 9.2.3.

[3] This finding, and the shape of this relationship, will also be discussed in more depth in the upcoming Section 9.2.1.

- Ensuring attention and research efforts are focused on 'key' areas of PAD space, as dictated by cluster best representatives;
- Classifying stimuli optimally to distinguish between different underlying distributions (in an effort to increase power);
- Reducing error / confounds when matching modalities as closely as possible, one stimulus at a time;
- Helping to find (rather than decide) the suitable number of stimulus groups for research;
- Parsimony;
- Flexibility;
- Reproducibility;
- Depending on cluster size, highlighting which areas of PAD space support heavier sampling of stimuli, and which are sparser and may not be as supportive of certain research designs / goals;
- Helping to identify ambiguous stimuli (i.e., stimuli with high membership uncertainty).

Testing words, sounds, images, and film clips

After creating matched triplets of affective words, sounds, and images, we tested these modalities on a new sample of participants, alongside one new modality: a group of 75 films. Of these, a subset of 25 films were selected based on how closely their PAD ratings matched the three other stimulus types.

This study also provided evidence that for images selected as best representatives from their respective clusters, PAD ratings were highly correlated between the original database norms and a local Edinburgh sample, despite any generational and cultural differences. Hence, this is a good indication that, in this case, our stimulus sampling method may be used directly on the database norms, without first needing to revalidate all norms locally. In fact, the similarity in PAD ratings between samples was such that clustering the stimuli using these new data led to a very similar classification to the one generated using the original database norms.

Other findings included the poor performance of the AffectButton relative to the Self-Assessment Manikin (SAM, Bradley & Lang, 1994). We did not find evidence to support the AffectButton's validity, given that for all stimulus modalities and for the dimensions of Arousal and Dominance, correlations were low between the SAM and the AffectButton. Only Valence was understood similarly between the two instruments. Hence, the AffectButton was discussed less from this point onwards, with validity concerns potentially related to the very small samples used in the original research involving this tool (Broekens & Brinkman, 2009, 2013).

In addition, only images and films were compared in terms of presence and negative physical symptoms (as comparing words and sounds with respect to presence would not make as much sense). We found that both high Valence and high Arousal boosted engagement / presence[4], and that film clips were more engaging than images, but as likely to lead to negative effects (i.e., there was no significant difference between films and images in terms of leading to negative symptom). Finally, intensely negative material was associated with more negative physical symptoms, whereas positive and serene material acted as a protective factor against such effects. Personal characteristics also influenced engagement results: males felt more present, and participants with higher scores on the Negative PANAS, as well as Far-East participants, were more susceptible to negative symptoms (suggesting the importance of controlling these factors in research).

Within logistic models and after keeping just the ⅓ most suitable films, we did not observe differences between any pair of clusters due to modality (as long as the SAM was used in measurement). Separately in linear mixed models predicting scores on each PAD dimension, we did not observe differences between the 4 stimulus modalities, with the exception of images which were more arousing and less Dominance-inspiring than words (again while assessed with SAM). Except for this, all modalities (including films) were non-significantly different from words, while taking into account a whole range of (participant-level) covariates. This suggests the method for selecting new stimulus modalities (i.e., testing three times more stimuli than ultimately required, only to pick the ⅓ most suitable, and then matching them) yielded adequate results overall.

Adding VEs without an HMD This study was a further investigation of affective stimulus modalities. We asked a new sample of participants to rate the ⅓ of films retained from the previous chapter, plus 75 new VEs - all tested without an HMD. Findings included good test-retest reliability for films (for this study compared to the previous one) - hence the previous strategy of selecting and matching stimuli led to useful, usable results.

However, we discovered that range restriction on PAD dimensions was too severe to attempt a matching process between the 75 new VEs introduced here, and previous stimulus modalities (i.e., words, sounds, images, films). VEs were not as negative as images or even sounds, nor were they as arousing as films. They also did not occupy areas of bivariate space where other stimulus modalities were still present (e.g., very high Valence, combined with very high Arousal). This range restriction issue casts further doubt on the idea that a "common emotional system" may exist, and provides more support for the idea that choosing one

[4] Given the factor structure we found for the ITC-SOPI-SF, the terms "presence" and "engagement" are used interchangeably here.

stimulus modality over another can bias results. This issue is also related to having identified only $k = 2$ clusters for VEs.

Intriguingly, the VEs used here were found to be *less* engaging than films, which might potentially be due to Second Life being unsuitable for the aims of this research, as will be discussed shortly. VEs were also associated with more negative physical symptoms than films.

Testing VEs with an HMD

Following the results obtained above, we were also interested to see whether increasing the immersiveness of VEs could alter and 'intensify' participant ratings. In order to test this, we asked a new sample of participants to explore 6 VEs using a head-mounted display (HMD).

We found that PAD ratings were not affected significantly by users wearing an HMD vs. navigating VEs displayed on a large monitor. However, HMD wearers felt higher levels of presence than large screen users - especially when exploring positive VEs. In contrast, negative VEs were associated with more negative physical symptoms and sickness. HMD wearers were not significantly more uncomfortable than monitor users (however it is also true that the HMD rating procedure was shorter than when a monitor was used).

Human vs. machine classification of VEs

This study was designed in order to investigate whether model-based clustering (MBC) can replicate the manner in which human participants classify emotional information. To this end, participants were asked to navigate a set of VEs, rate them on the PAD model, and classify them freely - with the liberty of leaving some VEs unclassified.

Multiple findings emerged from this study. Firstly, VE uncertainty (as computed using MBC) was found to influence the number of groups (i.e., category size) which participants classified the VEs into: the more uncertain / ambiguous the VE, the higher the chance of it being sorted into a smaller, fine-grained category. The opposite occurred for VEs with high certainty, which tended to be placed into larger, coarser categories. In addition, uncertainties also influenced the co-occurrence pattern between VEs pairs, with the most reliably occurring pairs in spontaneous participant classifications being those with near-zero uncertainties within their respective MBC clusters.

In addition, the MBC classification[5] was also linked to co-occurrence patterns from participant classifications: the majority of co-occurrences found in human classifications were established within the confines of the clusters identified with MBC - hence MBC was able to "retrieve" a large part of this pattern. This similarity could be a result of human and automatic classifications *both* relying

[5] Referring to way in which VEs were partitioned into clusters, regardless of uncertainty.

on PAD dimensions internally. Therefore, despite any assumed algorithmic differences, the two classifications (human-made and MBC) still bore considerable similarity, so that sampling emotional stimuli using cluster analysis for later use with human samples can be considered a reasonable course of action.

Another set of findings involves the Arousal dimension which (followed by Dominance), was the most important factor in determining both the "classifiability" of VEs, as well as their rank in terms of category prototypicality (although interestingly, MBC uncertainties were unrelated to the VE prototypicality hierarchies decided by human participants). This is further evidence that stimuli should not be sampled based on Valence alone.

However, participants may also be using other (semantic) dimensions rather than just the PAD model for their classifications, given that specific VEs co-occurred very strongly (e.g, all erotic stimuli). Also, humans often grossly overestimated the number of clusters identified by MBC, and hence do not appear to operate using parsimony criteria - another point of departure from MBC. Finally, a large amount of variability was also seen across individual classifications, which the current analysis (focusing on Valence, Arousal, Dominance, as well as the MBC classification and uncertainties) was not able to account for.

Real-life emotional experiences In this study, we asked participants to rate their emotional experiences at random times throughout the day using a phone app, over a period of two weeks. Participants submitted PAD ratings, and also classified their emotional episodes themselves. These data were then contrasted with other data previously collected in response to lab stimuli.

Three key types of findings were identified: firstly, concerning *the frequency* of various emotional phenomena in the real world, we discovered that more positive emotional events were discovered in real-life, relative to how lab stimuli are sampled and / or perceived by participants. In addition, the amount of Dominance-inducing emotional events was also higher in real life compared to the lab. In contrast, Arousal ratings did not differ between real-life and the lab environment. In terms of individual stimulus modalities, VEs differed from real-life only in terms of Valence ratings (with VEs being less positive than real-life, similarly to all other lab stimuli), but not Arousal or Dominance.

Secondly, *structural* differences in PAD space were also found: real-life events differed from lab stimuli both in terms of the richness of experiences (with lab stimuli showing restricted ranges on PAD compared to real-life events, and actually shown to be part of statistically different populations), as well as the types of bivariate relationships established between the PAD dimensions (with lab stimuli showing a marked quadratic trend, which was obviously different to the linear, orthogonal relationship found for real-life events).

Finally, the issue of neutrality emerged again, and was found to differ between real-life events (where neutrality represented the very centre of PAD space) and lab stimuli (collectively defined by medium Valence, but low Arousal, and low Dominance). This has implications for how baselines and control conditions should be chosen for research, which is discussed further in Section 9.2.2.

9.2 Main themes and contributions, and their relevance within the literature

9.2.1 The shape of PAD relationships, in particular between Valence and Arousal

Kuppens, Tuerlinckx, Russell, and Barrett (2013) discuss a total of 6 possible relationships which can be established between Valence and Arousal: **independence** (i.e., Valence and Arousal are orthogonal dimensions, with a relationship visualised as a straight horizontal line), **positive** or **negative linear relation** (i.e., an ascending or descending, straight diagonal line), **symmetrical quadratic relation** (i.e., a "U"-shaped curve where the halves defined by increasing Arousal are equidistant to medium Valence), **asymmetrical quadratic relation** (i.e., compared to the symmetrical relationship, here the increasing Arousal might show different slopes depending on whether Valence is above or below its midpoint; this option might also include a "positivity offset" or "a gap" in the relationship when Valence increases beyond a certain level (see the original work by Kuppens et al., 2013 for details), and finally, an **inverted quadratic relation** when Valence is a function of Arousal (i.e., an inverted "U"-shape, where affect is most pleasant at medium levels of Arousal are most pleasant, with extremes on either side becoming more unpleasant as they diverge from the optimal level).

Evidence suggesting one shape over another tends to place us within different theoretical frameworks concerning what Valence and Arousal ultimately represent (e.g., for quadratic trends, Arousal becomes an intensity measure / property of Valence, whereas for orthogonal models of Valence and Arousal, these two dimensions are separate entities in their own right). What is more interesting (or indeed, even concerning) is that the lab data collected in the current research always followed a quadratic trend (whether symmetrical or asymmetrical - see Figure 5.16, p. 260) between Valence and Arousal, whereas for real-world affect, these dimensions were found to be orthogonal instead. This is an indication that highly controlled / "artificial" testing conditions and stimuli might altogether mislead affective research, and suggest that Arousal is a property of Valence, when in fact it may be an altogether independent dimension.

Our results involving lab stimuli are confirmed by previous research repeatedly discovering quadratic trends between Valence and Arousal (when these are assessed using laboratory stimuli) - particularly as far as affective words are concerned (Bradley & Lang, 1999a; Redondo et al., 2007; Riegel et al., 2015; Soares, Comesaña, Pinheiro, Simões, & Frade, 2012; Warriner, Kuperman, & Brysbaert, 2013). Sounds (Bradley & Lang, 2007b), images (Bernat et al., 2006; Bradley, Codispoti, Cuthbert, & Lang, 2001; Bradley & Lang, 2007c; Constantinescu et al., 2016), and even the Velten technique (Jennings, McGinnis, Lovejoy, & Stirling, 2000) have also been found to exhibit this same trend between Arousal and Valence. In the light of such findings in this area, the orthogonal relationship we found between these Valence and Arousal within real-world data is particularly worthy of attention and further research.

However, other relationship shapes have also been found between these two dimensions, and it is difficult to pinpoint the source of these differences across stimulus types, as well as authors. For instance, Marchewka et al., 2014 found a linear correlation between Valence and Arousal within the NAPS image database. Null correlations (possible indicators of orthogonality) have also been found between Valence and Arousal (Barrett & Fossum, 2001; Kuppens, Van Mechelen, Nezlek, Dossche, & Timmermans, 2007; Yik et al., 1999).

As a tentative explanation, Kuppens et al. (2013) suggest that in some cases, the type of Valence - Arousal relationship is built into the measurement tool, e.g., the PANAS scales by Watson et al. (1988) include combinations of Valence and Arousal, so that these cannot emerge as orthogonal. This issue however does not afflict the SAM scales (Bradley & Lang, 1994) which assess each PAD dimension independently, and which were used consistently throughout this research - whether for lab stimuli or real-life emotional episodes. This being the case, the discrepancy between lab stimuli and real-life events still emerged, and is therefore particularly interesting - both theoretically and practically. It suggests that Valence and Arousal may indeed be orthogonal in natural environments, but that lab stimuli (at least those tested here) may only present a narrow window onto affective processing and the relationships established between these dimensions. Indeed, lab stimuli may be uniquely suited for representing only specific areas of the PAD space (conforming to the "U" shape), thus only giving the appearance of a quadratic trend between Valence and Arousal, when in fact outside the lab, the entire PAD space was easier to represent / sample from, over a period of only 2 weeks.

A particular difficulty with investigating the relationships between unobservable constructs such as Valence and Arousal, is that research regarding them tends to be conducted at the *nomothetic* level (i.e., with results averaged across persons / conditions), although the aim is usually to understand processes occurring at the *idiographic* level (i.e., within individuals). Sometimes these two perspectives can show major divergence, so that higher-level data cannot be used to draw direct inferences about all / any particular individual(s) (see Feldman, 1995; Kuppens et al., 2013). Consequently, such results become challenging to interpret, particularly given the variations found between individuals in terms of Valence - Arousal relationships (Barrett, 1998; Feldman, 1995; Kuppens, 2008; Kuppens et al., 2017).

We attempted to eschew this very issue and explore how well individual participants' ratings of the IAPS images were reflected by the averaged norms. Depending on each individual's PAD data, the well-known "U"-shaped Valence $\times$ Arousal relationship may potentially be a distortion produced by the averaging process. To investigate this, we requested the original data used to generate the IAPS norms (Bradley & Lang, 2007c), but this request remained unanswered by the authors (e-mail communication on November 6^{th}, 2015). Therefore, only average stimulus ratings were available to us, and in order for comparisons to make sense between the stimulus modalities discussed in this thesis, all data were aggregated at the nomothetic level - despite the risks discussed by Kuppens et al. (2013).

Finally, another alternative for explaining the differences between laboratory stimuli and real-life emotional episodes as far as the Valence - Arousal relationship is concerned, is that the PAD model (Mehrabian, 1995, 1996; Russell, 1980) itself may be an incomplete, or an otherwise poor approximation of how emotional stimuli are processed. Alternative (but similar) mathematical models have been proposed for capturing the structure of subjective affect (e.g., Mattek, Wolford, & Whalen, 2017), but further research will be required in this complex area.

9.2.2 The concept of emotional neutrality

Across this research, we used cluster analysis as a data-driven means to identify the group of stimuli most likely to function as "neutral" baseline. We found that due to the imperfect overlap between stimulus modalities, the meaning of "neutrality" (or where the cluster of neutral stimuli was located within PAD space) shifted across modalities. At this stage, it is not clear whether this means that perhaps the notion of "neutrality" varies by modality or research context, or rather if this is due to error variance. To be more precise, based on the IAPS, the closest cluster aligning to the idea of neutrality was defined by medium Valence, but low Arousal and high Dominance (for 'neutral', low-key and unchallenging content, such as the image of a woven basket). In contrast, for real-life emotional events, the clear candidate for neutrality was a cluster defined by the midpoint of all three PAD scales: Valence, as well as Arousal and Dominance.

In the wider literature, surprisingly little work appears to have been carried out to explicitly define neutrality, and determine the appropriate manner in which it should be measured. Instead, the idea and meaning of emotional neutrality is included implicitly within the debate for whether or not measurement scales (in particular for Valence) should be uni-/or bipolar. Depending on the school of thought and mixed empirical evidence, some authors advocate that positive and negative Valence should be separate, potentially even independent dimensions (Cacioppo & Berntson, 1994; Cacioppo, Gardner, & Berntson, 1997), whereas others view Valence as a continuum between positive and negative states, and hence would measure this construct using a bipolar scale (Bradley & Lang, 1994).

Depending on which model is selected, one may be able to account for different states: uni-polar scales, for instance, can measure ambivalent states (both positive and negative, to varying degrees), or 'emotion-less' states (i.e., where both positive and negative Valence are separately described by the minimum value on each scale) - assuming these even exist, since this has also been debated, with authors such as Russell and Barrett (1999) and Russell (2003) arguing that an emotional state ("core affect") is always present, whether or not individuals take note of it. On the other hand, bipolar scales (such as the SAM used here, Bradley & Lang, 1994) may be better suited to capture a different form of "neutrality", whenever individuals select the midpoint of the Valence scale.

Hence, the difference between these positions has implications for how "neutral states" are defined, and whether they may refer to the absence of emotion, or a balanced state halfway between positive and negative Valence, or even their coexistence within an ambivalent state. Tentatively though, fully neutral states seem to be different from ambivalent states (Gasper & Danube, 2016), or from emotional shock and numbness (Gallegos & Gasper, 2017), and surprisingly they also appear to require cognitive effort to be maintained (Gasper & Hackenbracht, 2015), suggesting this is a distinct emotional entity, rather than a purely 'affectless' state.

Clearly, distinguishing among all these possibilities could form the focus of an entire branch of research. However, based on the research presented here, it is important that researchers at the very least consider the issue of neutrality in more depth, and use multiple stimulus modalities as a default stance, in order to understand an emotional process in more detail, and protect against modality-specific bias. Meanwhile, the absence of a clear theoretical framework for defining neutrality (and its rare investigative pursuit) has deep practical consequences for experimental design: if there is no empirically-informed and / or theory-driven method for determining which stimuli are neutral and which are not, and what this means, then selecting appropriate control conditions and baselines for research becomes a delicate matter - one which could influence entire study results, and crucially, lead to unstable inferences.

9.2.3 The existence of a "central emotional system"

It has been proposed that any self-report, behaviour and / or physiological similarities identified between different stimulus modalities may provide information about how a "common emotional system" may be operating, in case this exists (Bradley & Lang, 2000; Redondo et al., 2008). On the other hand, due to the large variety of emotion elicitation methods (see section 1.3, p. 41), a different perspective is also possible - namely, that researchers should opt for an elicitation method appropriate for their study aims (Kappas & Descôteaux, 2004), chosen theoretical framework (e.g., discrete emotions or dimensional view), and desired range or intensity of emotional experience (Bujarski, Mischel, Dutton, Steele, & Cisler, 2015). Moreover, researchers must also "be aware of the assumptions underlying a given set of stimuli" (Keil & Miskovic, 2015). In fact, Grühn and Sharifian (2016) even proposed an "Emotion Matrix" to help with the decision process for selecting emotional stimuli - an implicit indication that they do not consider all affective stimuli to be equivalent in terms of utility and/or quality. The decision criteria proposed by these authors are: "ecological validity"[6], temporal resolution, controllability, complexity, and emotional intensity.

The data collected as part of the current research also did not support the idea that a central emotional system exists, operating similarly across stimulus modalities. This conclusion is based on the difficulties we encountered when attempting to match stimuli across modalities, or on the fact that "neutral" clusters varied by modality. Nonetheless, the surprising "U"-shaped relationship between Valence and Arousal was indeed common to all lab stimuli. However, this may indicate that all these lab modalities share the same form of bias (which is not present in real-world emotional events), rather than genuine similarity in terms of how they are processed.

Further examples for the differences between emotional stimulus modalities have been flagged by others: Uhrig et al. (2016) compared affective images and film clips, and surprisingly, found that images proved both more arousing and more useful for triggering the intended states than film clips, whereas somewhat confusingly, Bujarski et al. (2015); Westermann et al. (1996) consider films to be extremely effective precisely because they combine (dynamic) visual and auditory information. In a meta-analysis, Lench et al. (2011) compared various types of emotion elicitation methods: films, images, priming, music, Velten, imagery, text, behavioural techniques including the Directed Facial Action Task, social psychological methods, and autobiographical recall, and found that overall, images constituted the method associated with the largest effect sizes, surpassing even film clips - despite the latter being used more frequently in research, according to the authors. Perhaps an even better example is discussed by Bujarski

[6] As mentioned previously, "representative experimental design" is a more appropriate term (Brunswik, 1955).

et al. (2015) and involves arachnophobes averting their attention *away* from images of spiders, whereas in Virtual Reality environments, they tended to spend more time actually looking *towards* the spiders. For all these reasons, it would appear that a single, "central" emotional system is unable to account for all these findings, or at least no clearly articulated theoretical proposal has currently been advanced for such a system, which can accommodate/predict these diverse results.

Finally, according to Rouby, Fournel, and Bensafi (2016), information processing occurs more easily for congruent streams of multi-modal emotional information, and it is not yet clear what happens when sensory information is presented selectively to participants (i.e., only a visual information stream in the form of images, or auditory information only, in the case of emotional sounds etc.). Hence, it is either the case that all the differences between stimulus modalities are evidence against a modular central emotional system, or these differences may not even be relevant for the argument of whether or not such a central system exists, because they are so inconsistent with daily-life, multi-sensory experiences, that they are less meaningful and very prone to error.

9.2.4 The potential of VEs to elicit emotions

VR has been used successfully to elicit or modulate emotional states (see e.g., Baños et al., 2006; Chirico et al., 2016; Felnhofer et al., 2015; Serrano et al., 2013), and shows potential for use in research due to a unique combination of features: virtual stimuli may be both realistic and engaging enough to elicit 'genuine' affective reactions with good generalisability to the outside world (McMahan, 2003), while maintaining high degrees of experimental control and replicability (G. Young, 2010).

Arguably though, for VR to function as a proxy for real-life, and encourage participants to behave as they normally would outside testing conditions, a sense of "presence" must first be generated, i.e., a feeling of actually existing and operating fully within the VE (Baños et al., 2004; McMahan, 2003; Schubert et al., 2001). Emotions are usually considered an integral part of the process of achieving presence within a VE: for example, Riva et al. (2007) found the feeling of presence to be heightened within "emotional" VEs, but the directionality of this relationship remains unclear, i.e., whether perhaps it was the feeling of presence (acquired through some other means) which led to VEs being perceived as more emotional. Similarly, Freeman, Lessiter, Pugh, and Keogh (2005) also reported that presence ratings increased with happiness ratings, and Diemer et al. (2015) too concluded that overall, the stronger the emotional content associated with a VE, the greater the likelihood of finding an association between presence and emotion.

Freeman et al. (2005) go further to explain this relationship, and theorise that presence and Arousal, in particular, are related because presence is not only the perceived feeling of "being there" in a VE, but also the feeling of "doing there", therefore being able to react freely to stimuli. Hence, it is more likely for that presence can arise in response to emotional *and arousing* stimuli, which are a "*call to action*" on the part of the user.

These findings bear some similarity to those emerging from the current work: for instance, participants wearing an HMD did feel more present than large screen users, especially when exploring emotional, positive VEs - with presence assessed using a short form of the ITC-SOPI questionnaire (Lessiter et al., 2001). Arousal also contributed to a heightened feeling of presence, as anticipated by Freeman et al. (2005) above. An encouraging finding was that the boost in immersion (i.e., when wearing an HMD, as opposed to navigating VEs on a desktop monitor) did not come at the price of significantly more physical discomfort for participants (although it is also true that the HMD rating procedure was shorter than when a monitor was used). PAD ratings themselves were not affected by the use of an HMD vs. a large screen for the same VEs.

When comparing to other modalities, however, the VEs used here were found to be *less* engaging than films. VEs were also associated with more negative physical symptoms than films - both findings which make VR (or Second Life specifically) harder to recommend as research tools. Crucially, the VEs tested as part of this research were also associated with a far smaller range of emotional ratings on the PAD model (leading to $k = 2$), relative to other modalities. As to the reasons for these surprising findings, they may range from the manner in which presence was assessed using the ITC-SOPI-SF, to the particularities of Second Life, to equipment and technical shortcomings.

For instance, the original published work on the ITC-SOPI identified four factors as underlying participant responses: "Sense of Physical Space", "Engagement", "Ecological Validity", and "Negative Effects" (Lessiter et al., 2001), and we consulted with one of the authors of the ITC-SOPI on how to select items for a short form version which could represent each of these (personal communication with Jane Lessiter, on November 27, 2014). Despite this, in the current work we were only able to identify two ITC-SOPI dimensions: a single factor related to Presence / Engagement, and a second factor referring to Negative Effects / Negative Physical Symptoms.

This may be due to the existence of cross-loadings and correlations between the original four factors, or even the intervention of cultural and / or media-specific factors which may have exerted an unknown influence on results: the original ITC-SOPI questionnaire was tested on a sample of 604 participants, which varied greatly in terms of age (i.e., from 9 to 73 years) - and therefore probably also in experience with media forms such as VR. The same original sample also completed the ITC-SOPI in response to different forms of media, with all responses seemingly pooled together before assessing the dimensionality of the questionnaire. Despite this, no measurement invariance checks appear to have been conducted, e.g., between the various age groups, or between the multiple subsamples which used different forms of media. It is equally true that the sample sizes used in the current work were fairly small, and the version of the tool we tested was new (ITC-SOPI-SF) - so together all these factors may have contributed to the creation of a different factor structure, with an unknown impact on results.

Other factors have also been shown to affect VR procedures, and may have influenced the intensity / variability of the emotional responses we collected through this medium: quality of graphics and audio, personal relevance of the task (Houtkamp, 2012), a form of emotional blunting associated to gaming environments (Stuart, 2016), insufficient levels of interactivity which can damage the feeling of presence (given that "being there" has been linked to the ability to "do there", Ijsselsteijn, 2002), cultural differences and participants' level of education / familiarity with VR (Gorini et al., 2009), or even the use of Second Life itself - designed to cater to interests which may not overlap enough with academic research: self-therapy, a source of instant pleasures, liberation from social norms, a tool for self-expression, and exploration / novelty (Partala, 2011).

Furthermore, in terms of hardware options and the potential for inducing negative physical symptoms, poorer-quality VR can lead to suboptimal experiences due to: latency / delays in rendering time, head tracking inaccuracies, field of view limitations, image quality, and having to feel around for controls while wearing an HMD (Kim, Rosenthal, Zielinski, & Brady, 2014; Stein, 2016) - all of which were present here to varying degrees. The Oculus Rift DK1 HMD was used for this research, and has also been tied to negative symptoms (Davis, Nesbitt, & Nalivaiko, 2015; Tan, Leong, Shen, Dubravs, & Si, 2015) - indeed, even the following version of the Rift (DK2) was found to be similarly flawed (Hupont, Gracia, Sanagustín, & Gracia, 2015).

In this context, Pallavicini et al. (2013) also show that when faced with a choice between using VR "with breakdowns" [*sic*!] and using more traditional types of stimuli, it is more effective to resort to the latter, because the possibilities afforded and advantages offered by VR tend to vanish when the virtual experience is not smooth enough, and its capacity for inducing emotions is even lower than that provided by much cheaper media - a finding which certainly was supported by our own. Hence, the combination of Second Life and Oculus Rift DK1 may not have been ideally suited for the type of research outlined here, and custom-made VEs would probably represent a better alternative, to be paired with state-of-the-art HMDs (such as the HTC Vive).

Nonetheless, we believe that there are still strong reasons to investigate VR further (as discussed at the beginning of this section), particularly since other authors have found VR to be so convincing that adult participants can mistake a virtual limb (or even an opposite-gender/child's virtual body) as their own, and react very strongly when it is harmed, despite their own body being unaffected (see the research on the body transfer illusion, e.g., Banakou, Groten, & Slater, 2013; Kilteni, Normand, Sanchez-Vives, & Slater, 2012; Slater, Spanlang, Sanchez-Vives, & Blanke, 2010).

9.2.5 What is a suitable value for k?

Empirically, we found the 'appropriate' number of clusters / stimulus groups (k) to vary by modality, as well as classification method: model-based clustering (MBC), or spontaneous and free categorisations by human participants. We are not aware of any previous clear stance on the issue of how many stimulus groups should be used in research, from a perspective of experimental design - perhaps with the exception that the groups chosen to represent an independent variable should be as different as possible (regardless of their number), in order to boost power (Hallahan & Rosenthal, 1996). Theoretically though, this variation in k does present some problems for the discrete emotions view, which postulates the existence of a fixed number of automatic and universal emotion modules or programs (see our discussion in Section 1.1, p. 31).

In terms of MBC, we found $k = 5$ clusters within the IAPS database, 4 clusters within a matched and mixed-modality database containing IAPS images, ANEW words, and IADS-2 sounds, yet again 4 clusters when matched films were added to these three modalities, 2 clusters for VEs, and finally, 6 clusters when considering all these lab stimuli simultaneously. However, for real-life events, using MBC we only found $k = 3$ clusters. Since MBC does not allow to account for by-participant idiosyncrasies in the classification process (evidenced instead when human participants are allowed to freely classify emotional information - see Section 7.4.3, p. 330), in this case the variations in k across modalities are a direct consequence of all these modalities occupying different areas of PAD space - the reasons for which have been discussed above, relative to the existence (or absence) of a central emotional system.

For human categorisations of either lab stimuli (VEs), or real-life events (i.e., when participants grouped their own emotional events under self-made categories, within the phone app we designed), the values for k tended to be considerably higher: on average, participants structured real-life emotional events into approximately 9 categories (average across the entire sample), whereas for VEs, the sample grouped these into, on average, 7 categories (approximately). This difference between human classifiers and MBC (going from $k = 3$ to, on average, $k = 9$ for real-life emotional events, and from $k = 2$ to $k = 7$ on average for VEs) warrants further investigation, and suggests that human participants do not operate under parsimony constraints, unlike MBC (Fraley & Raftery, 2002, 2006). This may be due to human participants categorising/clustering information under a soft constraint to communicate more information via the classification, and therefore provide a more granular and detailed category structure, or due to a strategy for managing uncertainty (see following subsection).

Finally, as to the original question: "What is a suitable value for k?", a concise answer is that it depends. These findings only go so far as to suggest that a conscious and empirically-informed decision should be taken concerning how many stimulus groups should be used for research, rather than assuming a-priori that there should be just three, Valence-based stimulus groups, as is customary (see arguments made in Chapter 2, p. 65). However, further research will be required to explain the reasons behind the variations in k, in the hope that more concrete recommendations can be devised for appropriate and modality-independent stimulus sampling.

9.2.6 The classification process of emotional information

9.2.6.1 Parallels between MBC and human classifications of emotional information

Multiple parallels can be drawn between the manner in which humans categorise conceptual and/or emotional information, and the computational characteristics of MBC. Human participants classify information hierarchically, within categories with fuzzy borders, and around prototypical or basic-level concepts which are emblematic for the category in question (Edelstein & Shaver, 2007; Fehr & Russell, 1991; Rosch et al., 1976; Shaver et al., 1987). Similarly, in MBC, cases are assigned probabilistically to clusters (which therefore also present fuzzy borders). Via MBC, cases can also be sorted in terms of their (un)certainty of cluster membership, hence forming a hierarchical structure where cases with a high certainty of membership are akin to category prototypes / best representatives (Fraley & Raftery, 2002, 2006).

On the basis of these similarities between MBC and how humans categorise concepts/emotions, we investigated whether we can build an approximate model for how humans classify emotional information. To our knowledge, this has not been attempted before, and this parallel has led to a set of interesting findings: for instance, we found that higher MBC uncertainties translated into an inflation in the number of participant-made VE categories, which became finer-grained for such ambiguous material. Conversely, VEs with high certainty tended to be placed into larger, coarser categories.

This finding suggests that the results provided by MBC are indeed meaningful with regards to human performance, although it is unclear exactly what role statistical uncertainties play in this process. Potentially, and depending on their tolerance to ambiguity (Frenkel-Brunswik, 1949; Furnham & Marks, 2013; Furnham & Ribchester, 1995), participants may be engaging in an effortful uncertainty management/reduction strategy (see e.g., Tiedens & Linton, 2001) whereby they created a multitude of smaller categories to house uncertain items, so that these items may gain higher certainty than they otherwise would if placed in fewer, and coarser categories. The same strategy of creating more granular classes also presents the benefit of communicating more information on the category members within. This may explain why human participants did not follow parsimony criteria, and often overestimated the number of categories relative to MBC.

In addition, we also found that the $k = 2$ MBC solution "retrieved" a large part of the VE co-occurrence pattern seen in human classifications. A likely contributor to this finding is that both human and automatic classifications may rely on the same input information, despite potential algorithmic differences: the PAD dimensions, as these are considered to underlie emotional experiences in general (Mehrabian, 1996; Russell & Mehrabian, 1977). Such evidence of MBC validity when mimicking how humans classify emotional information can support the use of MBC-based stimulus classifications in research on human participants, as well as represent a tentative first step toward a more concrete theory/model of emotional categorisation.

Interestingly, the Arousal dimension distinguished itself as being the most important factor in determining both how "classifiable" the stimuli were, as well as their position within the hierarchy of human-made item prototypicality. In other words, it was easier to classify arousing items in absolute terms (rather than leave them unclassified), it was also easier to determine which category to place them in, and they also tended to be more representative / prototypical of their categories - more so than if these items merely presented high levels of certainty, as computed using MBC.

A potential explanation for these findings may relate to Arousal exerting an influence on memory - with more arousing information leaving longer lasting traces in memory (Kensinger & Corkin, 2004; LaBar & Phelps, 1998; Mather, 2007). If this is so, then any VEs remembered more easily post-exploration may have been more salient afterwards, during the emotional classification task (see also the discussion in Section 7.4.3, p. 330). Furthermore, these findings constitute additional evidence that stimuli should not be sampled based on Valence alone (see our arguments in Constantinescu et al., 2016), as the other PAD dimensions may also influence emotional processing, as is suggested here.

9.2.6.2 Research implications

Previous findings suggest it is often challenging to translate hierarchical (emotion) concepts / *discrete* types (e.g., fear, subdivided into terror, anxiety, panic etc.) into *continua* such as Valence and Arousal (Russell & Barrett, 1999). Indeed, when attempting to regress emotional categories onto PAD ratings or vice versa, these do not perfectly overlap, hence being considered complimentary perspectives (Stevenson et al., 2007). Due to this, some authors have opted to characterise affective stimulus databases in terms of both discrete (or basic) emotions and continuous affective dimensions, e.g, for the IAPS (Mikels, Fredrickson, et al., 2005), IADS (Stevenson & James, 2008), and affective word databases (Ferré, Guasch, Martínez-García, Fraga, & Hinojosa, 2017; Stadthagen-González, Ferré, Pérez-Sánchez, Imbault, & Hinojosa, 2017; Stevenson et al., 2007; Strauss & Allen, 2008).

And yet, despite the difficulties associated with translating one perspective into the other, experiencing emotions (described by such *continuous dimensions*) seems to be an integral part of *categorising* them (Brosch et al., 2010): for instance, Beatty et al. (2014) found that the same brain regions involved in experiencing positive vs. negative states were also recruited during an affective classification task. Furthermore, Niedenthal, Halberstadt, and Innes-Ker (1999) found that when participants experience various affective states, they are more likely to classify stimuli according to their emotional properties, rather than other non-emotional criteria. The human capacity of converting continuous affect into emotional categories appears to be spontaneous and effortless, so much so that some authors (Barrett, 2006; Russell & Barrett, 1999) suggest that categorising affect is an integral part of, or even synonymous with experiencing emotions.

In this context, our findings suggest that using cluster analysis algorithms such as MBC can deepen our understanding of how human participants translate continuous affect dimensions into emotional categories, particularly by using mixture models (as the basis of MBC) to distinguish between the likelihood of various affective states belonging to one emotional category over another. Mixture models also allow this classification to be fuzzy / probabilistic, in a way which is not possible when regressing emotional categories directly onto PAD ratings, or vice versa.

Another advantage for the framework used here was that participants were *not* forced to choose among a finite set of emotion labels (e.g., anger, fear, happiness etc.) under which to classify the affective stimuli, unlike the typical approach in basic emotion research (see a critique of these methods by e.g., DiGirolamo & Russell, 2016; Nelson & Russell, 2016). Such a forced-choice procedure might be an oversimplification of emotional classification processes occurring under naturalistic conditions. To our knowledge, a free classification task for emotional stimuli, unrelated to basic emotion categories, has not been attempted previously. Indeed, the goal of the research here was more ambitious - to study the categorisation process of emotional information more generally, rather than specifically with reference to basic emotions only. In doing so, this also avoided issues where participants would be tempted to rely too heavily on the category labels provided, instead of their own spontaneous responses.

However, our findings do not purport to fully explain how humans classify emotional information, with results indicating that other spontaneous influences are also present. For instance, one influence on emotional categorisation came from the part of semantic content, the contribution of which has been considered in the literature (see e.g., Beatty et al., 2014; Czekóová et al., 2016; Machajdik & Hanbury, 2010; Wang & He, 2008). For now, it is unclear how to quantify this influence, and how to judge its contribution to the categorisation process relative to other dimensions such as Arousal. Individual differences may also have affected the classification process: while Arousal was often a clear influence, its contribution varied widely by participant, perhaps based on the extent to which individuals may focus their attention on variations either in Valence or in Arousal levels (Barrett, 1998; Erbas, Ceulemans, Koval, & Kuppens, 2015; Feldman, 1995).

9.2.7 Validity of lab elicitation techniques vs. real-world emotional episodes

Laboratory research is particularly appealing due to the ability to carefully construct experimental tasks meant to control confounds and error variance (as much as possible). However, devising such research conditions may then lead to poor generalisability of results, given that tasks can become so artificial and restrictive that they no longer reflect how psychological processes operate in their natural ecology (Brunswik, 1955; Jerit et al., 2013; Parsons, 2015). Despite this danger, it appears that work carried out to verify the validity of experimental tasks against their real-world counterparts is seldom performed, and the idea of whether experimental results are still meaningful for the outside world tend to either be absent, or implicitly assumed without further empirical investigation.

Where validation work is indeed carried out to check how laboratory results generalise to naturalistic environments, conclusions can be a cause for concern. As an example, G. Anderson and Brown (1984) compared real and artificial casino environments, and found that measures of heart-rate, self-report and gambling behaviour were significantly different between conditions. In addition, Wilhelm and Grossman (2010); Wilhelm, Roth, and Sackner (2003) discuss the example of physiological ratings to stressful emotional stimuli in the lab, vs. when a participant was later watching her favourite soccer team on television. The authors explain that the difference in heart rate between the two stressful conditions was remarkable, and in the case of the real-life setting, was spontaneously maintained for over 1 hour.

Another area where laboratory research can be misleading, and diverge from results obtained in naturalistic settings, is ageing research. Isaacowitz and Stanley (2011) for instance discuss how older adults can appear to perform more poorly on "traditional" emotion recognition tests, relative to young adults. However, when implementing more 'ecologically valid' tests, age differences are attenuated. Similarly, differences have also been found between the laboratory and field research concerning emotion regulation strategies (Brans, Koval, Verduyn, Lim, & Kuppens, 2013). As a result, authors such as Trull and Ebner-Priemer (2013) are beginning to recommend that researchers combine experimental methods with field studies.

In this context, our research revealed a set of interesting findings. For example, data collected from participants during their daily lives was significantly more positive than any of the lab stimuli used - be it words, sounds, images, film clips, or VEs. This was unexpected, and suggests that the frequency of positive emotional episodes (or their weight) might be higher for healthy individuals in naturalistic environments, compared to how laboratory stimuli are routinely sampled (i.e., usually positive, negative and neutral[7] stimuli *in equal amounts*). Findings by Diener and Diener (1996) and Zelenski and Larsen (2000) similarly confirmed the widespread nature of positive affect in daily life.

In light of such results, current research findings could be recontextualised. For instance, it would be difficult to completely exclude an alternative interpretation of the negativity effect / bias (i.e., the tendency of (young) adults to attend to, learn from, or otherwise process negative information preferentially instead of positive information, possibly for evolutionary and self-preservation reasons, Vaish, Grossmann, & Woodward, 2008): if negative emotional episodes simply seem to occur less frequently in the real world, then perhaps novelty could be why these are attended to more closely in research labs. Importantly, it is difficult to discern between such possibilities unless researchers start extending their measurement models and designs to incorporate field studies as well.

Another reason to strongly consider carrying out field research (in addition to any laboratory experiments) relates to range restriction and ceiling effects. Our results suggest (similarly to Wilhelm & Grossman, 2010) that emotional tasks in labs can obscure the true range of emotional responses which participants are capable of. In doing so, various areas of PAD space can selectively remain unpopulated, and thus lead to bias and error for instance when assessing the relationship between Valence and Arousal (see Section 9.2.1, p. 380).

[7] Depending on the definition of neutrality - see discussion in previous pages.

9.3 Limitations

Firstly, the work discussed here has relied exclusively on self-report data, which is only one facet of emotions (Coan & Allen, 2007). Ideally, other components should also be measured to gain a more profound understanding of emotional processes, e.g., behavioural, physiological and neurological measures.

Secondly, the sample sizes used throughout this work have tended to be relatively small, particularly for the phone app study. In this case, however, repeated measures were taken which might compensate for the smaller number of participants. Also related to sampling, participants originated from a variety of cultural backgrounds, which may have exerted an influence on results. Our phone app was also only suitable for Android users (vs. iOS, due to objective reasons), which may have introduced a certain level of bias into results. However, given the extremely wide popularity of this OS, it is unlikely that Android and e.g., Apple users are different in striking ways.

Finally, all VEs used in this research were selected from Second Life, which is not specifically designed for research, and also required that any potential changes made to the environments be tracked by the researcher and used as covariates. Ideally, the virtual stimuli used should have been custom-made, and remained fixed for the duration of the research, but this was not feasible.

9.4 Future directions

As mentioned above, the research carried out here relied heavily on self-report measures. However, Mauss and Robinson (2009) suggest that emotions should ideally be measured from multiple angles jointly: behavioural and physiological, alongside self-reports. This could be attempted in the future, and perhaps such additional dimensions (e.g., physiological measures) could be factored into MBC solutions, for higher discrimination between clusters. This multi-pronged approach might also help with the investigation of complex issues such as emotional neutrality.

We would also like to invite or encourage replication work for the main findings discussed here, e.g., the differences between the shape of the bivariate relationship between Valence and Arousal, when this is assessed for lab stimuli vs. when the points refer to real-life emotional episodes, or the role of Arousal in human categorisation processes. It is also important to replicate these findings under new conditions, e.g., when using other stimuli such as film clips instead of VEs, and then reassessing whether Arousal still plays an important role in how participants classify this emotional information. "Category drift" over time is another process worth investigating as an extension to the current work, i.e., how participants build, and perhaps revise, their emotional categories over time.

Other future improvements could also include the use of a more powerful, within-subjects design (with the same participants rating both lab stimuli and real-life emotions), as well as extending the stimulus matching process to include further stimulus modalities: e.g., emotional music, Velten, facial expressions, autobiographic recall etc. Ideally, these participants should also be sampled from the same culture, in order to reduce confounds.

Finally, for the future it would be essential to reassess the potential of VR to elicit emotions, but with better-quality equipment (e.g., most up-to-date graphics and HMD), and using custom built VEs which reflect specific desirable properties for research (unlike Second Life). Indeed, if these changes are shown to reduce the differences between VR stimuli and real-life emotion data, then a highly useful development would be to create a database of freely available VR stimuli / environments.

9.5 Conclusions

Based on all the findings discussed as part of this doctoral research, a few highlights and recommendations are worth mentioning: for instance, sampling research stimuli "manually" is an inefficient strategy, which can conveniently be replaced by cluster analysis algorithms. These data-driven methods can easily create partitions based on the specific (database of) affective stimuli intended for use. Moreover, clustering algorithms can achieve this by taking into account all three PAD dimensions, rather than just e.g., Valence - a highly desirable feature, as we found Arousal to play a role in how participants structure emotional information hierarchically, and Dominance to reveal outliers where the other two PAD dimensions failed. Depending on the characteristics of a given stimulus set, cluster analysis can also identify the group of stimuli which is closest to "neutrality". However, the empirical role of cluster analysis in identifying such suitable controls/baselines will need to be doubled by a better theoretical understanding of emotional neutrality in the future.

In addition, the current research highlighted the importance of carrying out field studies, as a yardstick against which to validate results from controlled environments. We found that the sole use of lab stimuli can bias both statistical model results, as well as the shape of multivariate distributions, both of which may change the theoretical significance of findings. In fact, none of the five stimulus modalities tested here proved to yield data which is particularly similar to real-life - including VR. However, in order to gain more definitive answers, further research will be needed surrounding virtual stimuli, especially if it is possible to employ custom-made VEs and advanced hardware.

This doctoral research has also made several methodological contributions. For instance, we proposed a general framework for gathering mixed-modality stimulus sets (i.e., images, words, sounds and films), classifying them into groups, and then selecting the most prototypical/representative stimuli from each group for experimental use. In addition, in our investigation of emotional categorisation, we used a novel research design (allowing participants to create any number of emotional categories they desired, as well as leave items unclassified), and a novel approach for data analysis, centred around the use of cluster analysis as a model for emotional categorisation. Finally, we also created an Android phone app to collect affective ratings for real-life emotional episodes, which is available from the author's GitHub account (see relevant chapter).

Part VII

Bibliography

Anderson, E., Siegel, E. H., & Barrett, L. F. (2011). What you feel influences what you see: The role of affective feelings in resolving binocular rivalry. *Journal of Experimental Social Psychology*, *47*(4), 856–860.

Anderson, G., & Brown, R. I. F. (1984). Real and laboratory gambling, sensation-seeking and arousal. *British Journal of Psychology*, *75*(3), 401–410.

Araujo, D., Davids, K., & Passos, P. (2007). Ecological validity, representative design, and correspondence between experimental task constraints and behavioral setting: Comment on Rogers, Kadar, and Costall (2005). *Ecological Psychology*, *19*(1), 69–78.

Arnold, M. B. (1960). *Emotion and personality.* New York: Columbia University Press.

Averill, J. (1980). A constructionist view of emotion. In R. Plutchik & H. Kellerman (Eds.), *Emotion: Theory, Research, and Experience* (Vol. 1, pp. 305–340). New York: Academic Press.

Averill, J. R. (1983). Studies on anger and aggression: Implications for theories of emotion. *American Psychologist*, *38*(11), 1145.

Averill, J. R., & Nunley, E. P. (1988). Grief as an Emotion and as a Disease: A Social-Constructionist Perspective. *Journal of Social Issues*, *44*(3), 79–95.

Bagby, R. M., Ayearst, L. E., Morariu, R. A., Watters, C., & Taylor, G. J. (2014). The Internet administration version of the 20-item Toronto Alexithymia Scale. *Psychological Assessment*, *26*(1), 16.

Bagby, R. M., Parker, J. D., & Taylor, G. J. (1994). The twenty-item Toronto Alexithymia Scale—I. Item selection and cross-validation of the factor structure. *Journal of Psychosomatic Research*, *38*(1), 23–32.

Bagby, R. M., Taylor, G. J., & Parker, J. D. (1994). The twenty-item Toronto Alexithymia Scale—II. Convergent, discriminant, and concurrent validity. *Journal of Psychosomatic Research*, *38*(1), 33–40.

Ball, G. H., & Hall, D. J. (1965). *ISODATA, a novel method of data analysis and pattern classification (Technical report)* (Tech. Rep.). Menlo Park, CA: Stanford Research Institute.

Banakou, D., Groten, R., & Slater, M. (2013). Illusory ownership of a virtual child body causes overestimation of object sizes and implicit attitude changes. *Proceedings of the National Academy of Sciences*, *110*(31), 12846–12851.

Baños, R. M., Botella, C., Alcañiz, M., Liaño, V., Guerrero, B., & Rey, B. (2004). Immersion and emotion: their impact on the sense of presence. *CyberPsychology & Behavior*, *7*(6), 734–741.

Baños, R. M., Liaño, V., Botella, C., Alcañiz, M., Guerrero, B., & Rey, B. (2006). Changing induced moods via virtual reality. In *International Conference on Persuasive Technology* (pp. 7–15).

Bard, P. (1928). A diencephalic mechanism for the expression of rage with special reference to the sympathetic nervous system. *American Journal of Physiology*, *84*, 490–513.

Barrett, L. F. (1998). Discrete emotions or dimensions? The role of valence focus and arousal focus. *Cognition & Emotion*, *12*(4), 579–599.

Barrett, L. F. (2006). Solving the emotion paradox: Categorization and the experience of emotion. *Personality and Social Psychology Review*, *10*(1), 20–46.

Barrett, L. F., & Fossum, T. (2001). Mental representations of affect knowledge. *Cognition & Emotion*, *15*(3), 333–363.

Barsade, S. G. (2002). The ripple effect: Emotional contagion and its influence on group behavior. *Administrative Science Quarterly*, *47*(4), 644–675.

Bartolini, E. E. (2011). *Eliciting emotion with film: Development of a stimulus set* (Unpublished doctoral dissertation). Wesleyan University.

Barton, K. (2016). *Multi-model inference. R package version 1.15.6.*

Bates, D., Maechler, M., Bolker, B., Walker, S., et al. (2014). lme4: Linear mixed-effects models using Eigen and S4. *R package version*, *1*(7).

Beatty, E. L., Vartanian, O., Muller-Gass, A., Robertson, J. A., Mandel, D. R., & Stergiopoulos, S. (2014). Neuroanatomical correlates of categorizing emotional valence. *Neuroreport*, *25*(11), 854–859.

Bechara, A., & Damasio, A. R. (2005). The somatic marker hypothesis: A neural theory of economic decision. *Games and Economic Behavior*, *52*(2), 336–372.

Belin, P., Zatorre, R. J., Lafaille, P., Ahad, P., & Pike, B. (2000). Voice-selective areas in human auditory cortex. *Nature*, *403*(6767), 309–312.

Bernat, E., Patrick, C. J., Benning, S. D., & Tellegen, A. (2006). Effects of picture content and intensity on affective physiological response. *Psychophysiology*, *43*(1), 93–103.

Bond, C. V. A. (1998). Personal relevance is an important dimension for visceral reactivity in emotional imagery. *Cognition & Emotion*, *12*(2), 231–242.

Borcard, D., Gillet, F., & Legendre, P. (2011). *Numerical ecology with R.* New York: Springer.

Bouchard, S. (2010). Emotions and the Emotional Valence Afforded by the Virtual Environment. In J.-J. Kim (Ed.), *Virtual Reality* (pp. 501–514). InTech. Online.

Bradley, M. M., Codispoti, M., Cuthbert, B. N., & Lang, P. J. (2001). Emotion and motivation I: defensive and appetitive reactions in picture processing. *Emotion*, *1*(3), 276.

Bradley, M. M., & Lang, P. J. (1994). Measuring emotion: the Self-Assessment Manikin and the semantic differential. *Journal of Behavior Therapy and Experimental Psychiatry*, *25*(1), 49–59.

Bradley, M. M., & Lang, P. J. (1999a). *Affective norms for English words (ANEW): Instruction manual and affective ratings* (Tech. Rep.). Technical Report C-1, The Center for Research in Psychophysiology, University of Florida.

Bradley, M. M., & Lang, P. J. (1999b). *The International affective digitized sounds (IADS): stimuli, instruction manual and affective ratings.* NIMH Center for the Study of Emotion and Attention.

Bradley, M. M., & Lang, P. J. (2000). Affective reactions to acoustic stimuli. *Psychophysiology*, *37*(2), 204–215.

Bradley, M. M., & Lang, P. J. (2007a). Affective Norms for English Text (ANET): Affective ratings of text and instruction manual. *Techical Report. D-1, University of Florida, Gainesville, FL.*

Bradley, M. M., & Lang, P. J. (2007b). The International Affective Digitized Sounds (IADS-2): Affective ratings of sounds and instruction manual. *University of Florida, Gainesville, FL, Tech. Rep. B-3*.

Bradley, M. M., & Lang, P. J. (2007c). The International Affective Picture System (IAPS) in the Study of Emotion and Attention. In J. A. Coan & J. Allen (Eds.), *Handbook of Emotion Elicitation and Assessment.* (pp. 29–46). New York: Oxford University Press.

Bradley, M. M., & Lang, P. J. (2010). *Affective Norms for English Words (ANEW): Instruction manual and affective ratings.* (Tech. Rep.). Technical Report C-2, University of Florida, Gainesville, FL.

Brans, K., Koval, P., Verduyn, P., Lim, Y. L., & Kuppens, P. (2013). The regulation of negative and positive affect in daily life. *Emotion*, *13*(5), 926.

Brehmer, B., & Dörner, D. (1993). Experiments with computer-simulated microworlds: Escaping both the narrow straits of the laboratory and the deep blue sea of the field study. *Computers in Human Behavior*, *9*(2-3), 171–184.

Brewer, D., Doughtie, E. B., & Lubin, B. (1980). Induction of mood and mood shift. *Journal of Clinical Psychology*, *36*(1), 215–226.

Briesemeister, B. B., Kuchinke, L., & Jacobs, A. M. (2011). Discrete emotion norms for nouns: Berlin affective word list (DENN–BAWL). *Behavior Research Methods*, *43*(2), 441.

Brock, R. L., Barry, R. A., Lawrence, E., Dey, J., & Rolffs, J. (2012). Internet Administration of Paper-and-Pencil Questionnaires Used in Couple Research Assessing Psychometric Equivalence. *Assessment*, *19*(2), 226–242.

Broekens, J., & Brinkman, W.-P. (2009). Affectbutton: Towards a standard for dynamic affective user feedback. In *Affective Computing and Intelligent Interaction and Workshops, 2009. ACII 2009. 3rd International Conference on* (pp. 1–8).

Broekens, J., & Brinkman, W.-P. (2013). AffectButton: A method for reliable and valid affective self-report. *International Journal of Human-Computer Studies*, *71*(6), 641–667.

Brosch, T., Pourtois, G., & Sander, D. (2010). The perception and categorisation of emotional stimuli: A review. *Cognition and Emotion*, *24*(3), 377–400.

Brunswik, E. (1955). Representative design and probabilistic theory in a functional psychology. *Psychological Review*, *62*(3), 193–217.

Bujarski, S. J., Mischel, E., Dutton, C., Steele, J. S., & Cisler, J. (2015). The Elicitation and Assessment of Emotional Responding. In K. Babson & M. Feldner (Eds.), *Sleep and Affect: Assessment, Theory, and Clinical Implications* (pp. 91–118). Amsterdam: Academic Press.

Burton, K. W., & Kaszniak, A. W. (2006). Emotional experience and facial expression in Alzheimer's disease. *Aging, Neuropsychology, and Cognition*, *13*(3-4), 636–651.

Cacioppo, J. T., & Berntson, G. G. (1994). Relationship between attitudes and evaluative space: A critical review, with emphasis on the separability of positive and negative substrates. *Psychological Bulletin*, *115*(3), 401.

Cacioppo, J. T., Berntson, G. G., Larsen, J. T., Poehlmann, K. M., Ito, T. A., et al. (2000). The psychophysiology of emotion. In M. Lewis & J. Haviland-Jones (Eds.), *Handbook of Emotions* (Vol. 2, pp. 173–191). New York: Guilford Press.

Cacioppo, J. T., Gardner, W. L., & Berntson, G. G. (1997). Beyond bipolar conceptualizations and measures: The case of attitudes and evaluative space. *Personality and Social Psychology Review*, *1*(1), 3–25.

Cahill, L., & McGaugh, J. L. (1995). A novel demonstration of enhanced memory associated with emotional arousal. *Consciousness and Cognition*, *4*(4), 410–421.

Caliński, T., & Harabasz, J. (1974). A dendrite method for cluster analysis. *Communications in Statistics - Theory and Methods*, *3*(1), 1–27.

Cannon, W. B. (1927). The James-Lange theory of emotions: A critical examination and an alternative theory. *The American Journal of Psychology*, *39*(1/4), 106–124.

Cantor, J. R., Zillmann, D., & Bryant, J. (1975). Enhancement of experienced sexual arousal in response to erotic stimuli through misattribution of unrelated residual excitation. *Journal of Personality and Social Psychology*, *32*(1), 69.

Carvalho, S., Leite, J., Galdo-Álvarez, S., & Gonçalves, Ó. F. (2012). The emotional movie database (EMDB): A self-report and psychophysiological study. *Applied Psychophysiology and Biofeedback*, *37*(4), 279–294.

Chirico, A., Yaden, D. B., Riva, G., & Gaggioli, A. (2016). The potential of virtual reality for the investigation of awe. *Frontiers in Psychology*, *7*, 1766.

Christie, I. C., & Friedman, B. H. (2004). Autonomic specificity of discrete emotion and dimensions of affective space: A multivariate approach. *International Journal of Psychophysiology*, *51*(2), 143–153.

Coan, J. A., & Allen, J. J. (2007). *Handbook of emotion elicitation and assessment.* New York: Oxford University Press.

Constantinescu, A. C., Wolters, M., Moore, A., & MacPherson, S. E. (2016). A cluster-based approach to selecting representative stimuli from the International Affective Picture System (IAPS) database. *Behavior Research Methods*, 1–17.

Coppin, G., & Sander, D. (2010). Théories et concepts contemporains en psychologie de l'émotion. In C. Pelachaud (Ed.), *Systèmes d'Interaction Émotionnelle* (pp. 25–56). Paris: Hermès science publications-Lavoisier.

Cornelius, R. R. (2000). Theoretical approaches to emotion. In *ISCA Tutorial and Research Workshop (ITRW) on Speech and Emotion.*

Crivelli, C., Jarillo, S., Russell, J. A., & Fernández-Dols, J.-M. (2016). Reading emotions from faces in two indigenous societies. *Journal of Experimental Psychology: General*, *145*(7), 830.

Crivelli, C., Russell, J. A., Jarillo, S., & Fernández-Dols, J.-M. (2017). Recognizing Spontaneous Facial Expressions of Emotion in a Small-Scale Society of Papua New Guinea. *Emotion*, *17*(2), 337–347.

Csikszentmihalyi, M., & Larson, R. (2014). Validity and reliability of the experience-sampling method. In M. Csikszentmihalyi (Ed.), *Flow and the Foundations of Positive Psychology: The Collected Works of Mihaly Csikszentmihalyi* (pp. 35–54). Springer.

Csikszentmihalyi, M., Larson, R., & Prescott, S. (1977). The ecology of adolescent activity and experience. *Journal of Youth and Adolescence*, *6*(3), 281–294.

Czekóová, K., Shaw, D. J., Urbánek, T., Chládek, J., Lamoš, M., Roman, R., & Brázdil, M. (2016). What's the meaning of this? A behavioral and neurophysiological investigation into the principles behind the classification of visual emotional stimuli. *Psychophysiology*, *53*(8), 1203–1216.

Damasio, A. R., Everitt, B. J., & Bishop, D. (1996). The somatic marker hypothesis and the possible functions of the prefrontal cortex. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *351*(1346), 1413–1420.

Darwin, C. (2005). *The expression of the emotions in man and animals.* New York, NY: Appelton. (Original work published 1872).

Davis, S., Nesbitt, K., & Nalivaiko, E. (2015). Comparing the onset of cybersickness using the Oculus Rift and two virtual roller coasters. In *Proceedings of the 11th Australasian Conference on Interactive Entertainment (IE 2015)* (Vol. 27, p. 30).

De Graaf, T., & Sack, A. (2011). Null results in TMS: from absence of evidence to evidence of absence. *Neuroscience & Biobehavioral Reviews*, *35*(3), 871–877.

de Arcos, F. A., Verdejo-García, A., Peralta-Ramírez, M. I., Sánchez-Barrera, M., & Pérez-García, M. (2005). Experience of emotions in substance abusers exposed to images containing neutral, positive, and negative affective stimuli. *Drug and Alcohol Dependence*, *78*(2), 159–167.

Deen, B., Pitskel, N. B., & Pelphrey, K. A. (2011). Three systems of insular functional connectivity identified with cluster analysis. *Cerebral Cortex*, *21*(7), 1498–1506.

Delplanque, S., N'diaye, K., Scherer, K., & Grandjean, D. (2007). Spatial frequencies or emotional effects?: A systematic measure of spatial frequencies for IAPS pictures by a discrete wavelet analysis. *Journal of Neuroscience Mthods*, *165*(1), 144–150.

Dhami, M. K., Hertwig, R., & Hoffrage, U. (2004). The role of representative design in an ecological approach to cognition. *Psychological Bulletin*, *130*(6), 959.

Diemer, J., Alpers, G. W., Peperkorn, H. M., Shiban, Y., & Mühlberger, A. (2015). The impact of perception and presence on emotional reactions: a review of research in virtual reality. *Frontiers in Psychology*, *6*, 26.

Diener, E., & Diener, C. (1996). Most people are happy. *Psychological Science*, *7*(3), 181–185.

DiGirolamo, M. A., & Russell, J. A. (2016). The Emotion Seen in a Face Can Be a Methodological Artifact: The Process of Elimination Hypothesis. *Emotion*, No Pagination Specified.

Dimitriadou, E., Dolničar, S., & Weingessel, A. (2002). An examination of indexes for determining the number of clusters in binary data sets. *Psychometrika*, *67*(1), 137–159.

Dinno, A. (2009). Exploring the sensitivity of Horn's parallel analysis to the distributional form of random data. *Multivariate Behavioral Research*, *44*(3), 362–388.

Dixon, T. (2003). *From passions to emotions: The creation of a secular psychological category.* Cambridge: Cambridge University Press.

Do, C. B., & Batzoglou, S. (2008). What is the expectation maximization algorithm? *Nature Biotechnology*, *26*(8), 897–900.

Dolnicar, S., Grabler, K., & Mazanec, J. A. (1999). A tale of three cities: Perceptual charting for analysing destination images. In A. G. Woodside, G. I. Crouch, J. A. Mazanec, M. Opperman, & M. Y. Sakai (Eds.), *Consumer Psychology of Tourism, Hospitality and Leisure* (pp. 39–62). Wallingford, UK: CAB International Publishing.

Dunlosky, J., Bottiroli, S., & Hartwig, M. (2009). Sins Committed in the Name of Ecological Validity. In D. J. Hacker, J. Dunlosky, & A. C. Graesser (Eds.), *Handbook of Metacognition in Education* (pp. 430–439). New York: Routledge.

Edelstein, R. S., & Shaver, P. R. (2007). A cross-cultural examination of lexical studies of self-conscious emotions. In J. L. Tracy, R. W. Robins, & J. P. Tangney (Eds.), *The Self-Conscious Emotions* (pp. 194–208). New York, NY: The Guilford Press.

Eerola, T., & Vuoskoski, J. K. (2010). A comparison of the discrete and dimensional models of emotion in music. *Psychology of Music*.

Eizenman, M., Lawrence, H. Y., Grupp, L., Eizenman, E., Ellenbogen, M., Gemar, M., & Levitan, R. D. (2003). A naturalistic visual scanning approach to assess selective attention in major depressive disorder. *Psychiatry Research*, *118*(2), 117–128.

Ekman, P. (2009). Expression and the nature of emotion. In K. R. Scherer & P. Ekman (Eds.), *Approaches to Emotion* (pp. 319–344). New York: Psychology Press, Taylor & Francis Group.

Ekman, P., Friesen, W., & Ellsworth, P. (1982). What emotion categories or dimensions can observers judge from facial behaviour? In, P. Ekman. *Emotion in the Human Face*, 39–55.

Ekman, P., & Friesen, W. V. (1971). Constants across cultures in the face and emotion. *Journal of Personality and Social Psychology*, *17*(2), 124.

Ekman, P., & Friesen, W. V. (1986). A new pan-cultural facial expression of emotion. *Motivation and Emotion*, *10*(2), 159–168.

Ekman, P., Levenson, R. W., & Friesen, W. V. (1983). Autonomic nervous system activity distinguishes among emotions. *Science*, *221*(4616), 1208–1210.

Engmann, S., & Cousineau, D. (2011). Comparing distributions: the two-sample Anderson-Darling test as an alternative to the Kolmogorov-Smirnoff test. *Journal of Applied Quantitative Methods*, *6*(3), 1–17.

Erbas, Y., Ceulemans, E., Koval, P., & Kuppens, P. (2015). The role of valence focus and appraisal overlap in emotion differentiation. *Emotion*, *15*(3), 373.

Estes, Z., & Adelman, J. S. (2008). Automatic vigilance for negative words is categorical and general. *Emotion*, *8*(4), 453–457.

Fann, J. R., Berry, D. L., Wolpin, S., Austin-Seymour, M., Bush, N., Halpenny, B., ... McCorkle, R. (2009). Depression screening using the Patient Health Questionnaire-9 administered on a touch screen computer. *Psycho-Oncology*, *18*(1), 14–22.

Fehr, B., & Russell, J. A. (1991). The concept of love viewed from a prototype perspective. *Journal of Personality and Social Psychology*, *60*(3), 425.

Feldman, L. A. (1995). Valence focus and arousal focus: Individual differences in the structure of affective experience. *Journal of Personality and Social Psychology*, *69*(1), 153.

Feldman Barrett, L., & Russell, J. A. (1998). Independence and bipolarity in the structure of current affect. *Journal of Personality and Social Psychology*, *74*(4), 967.

Felnhofer, A., Kothgassner, O. D., Schmidt, M., Heinzle, A.-K., Beutl, L., Hlavacs, H., & Kryspin-Exner, I. (2015). Is virtual reality emotionally arousing? Investigating five emotion inducing virtual park scenarios. *International Journal of Human-Computer Studies*, *82*, 48–56.

Ferguson, C. J., & Heene, M. (2012). A vast graveyard of undead theories: Publication bias and psychological science's aversion to the null. *Perspectives on Psychological Science*, *7*(6), 555–561.

Ferré, P., Guasch, M., Martínez-García, N., Fraga, I., & Hinojosa, J. A. (2017). Moved by words: Affective ratings for a set of 2,266 Spanish words in five discrete emotion categories. *Behavior Research Methods*, *49*(3), 1082–1094.

Field, A., Miles, J., & Field, Z. (2012). *Discovering statistics using R*. London: Sage publications.

Fontaine, J. (2009). Dimensional emotion models. In D. Sander & K. Scherer (Eds.), *Oxford Companion to Emotion and Affective Sciences* (pp. 119–120). Oxford: Oxford University Press.

Fox, J., Arena, D., & Bailenson, J. N. (2009). Virtual reality: A survival guide for the social scientist. *Journal of Media Psychology*, *21*(3), 95–113.

Fraley, C., & Raftery, A. E. (2002). Model-based clustering, discriminant analysis, and density estimation. *Journal of the American Statistical Association*, *97*(458), 611–631.

Fraley, C., & Raftery, A. E. (2006). *MCLUST version 3: An R package for normal mixture modeling and model-based clustering (Report No. 504)* (Tech. Rep.). Seattle, WA: University of Washington, Dept. of Statistics.

Fredrickson, M. M. (2010). *Speeding up Optmatch while improving match quality [Blog post].* (Retrieved on September 17, 2015 from: http://www.markmfredrickson.com/thoughts/2010-11-12-speeding-up-optmatch-while-improving-match-quality.html)

Freeman, J., Lessiter, J., Pugh, K., & Keogh, E. (2005). When presence and emotion are related, and when they are not. In *The 8th Annual International Workshop on Presence* (pp. 21–23).

Frenkel-Brunswik, E. (1949). Intolerance of ambiguity as an emotional and perceptual personality variable. *Journal of Personality*, *18*(1), 108–143.

Frijda, N. H. (1986). *The emotions: Studies in emotion and social interaction.* Cambridge: Cambridge University Press.

Furnham, A., & Marks, J. (2013). Tolerance of ambiguity: A review of the recent literature. *Psychology*, *4*(09), 717.

Furnham, A., & Ribchester, T. (1995). Tolerance of ambiguity: A review of the concept, its measurement and applications. *Current psychology*, *14*(3), 179–199.

Gallegos, J. M., & Gasper, K. (2017). Differential Effects of Rejection and Acceptance on Feeling Shocked, Numb, and Neutral.

Gasper, K., & Danube, C. L. (2016). The Scope of Our Affective Influences: When and How Naturally Occurring Positive, Negative, and Neutral Affects Alter Judgment. *Personality and Social Psychology Bulletin*, *42*(3), 385–399.

Gasper, K., & Hackenbracht, J. (2015). Too busy to feel neutral: Reducing cognitive resources attenuates neutral affective states. *Motivation and Emotion*, *39*(3), 458–466.

Gelman, A., Hill, J., & Yajima, M. (2012). Why we (usually) don't have to worry about multiple comparisons. *Journal of Research on Educational Effectiveness*, *5*(2), 189–211.

Gilet, A. (2008). Procédures d'induction d'humeurs en laboratoire: une revue critique [Mood induction procedures: A critical review]. *L'encéphale*, *34*, 233–239.

Gillath, O., McCall, C., Shaver, P. R., & Blascovich, J. (2008). What can virtual reality teach us about prosocial tendencies in real and virtual environments? *Media Psychology*, *11*(2), 259–282.

Glenn, C. R., Blumenthal, T. D., Klonsky, E. D., & Hajcak, G. (2011). Emotional reactivity in nonsuicidal self-injury: Divergence between self-report and startle measures. *International Journal of Psychophysiology*, *80*(2), 166–170.

Gomez, P., & Danuser, B. (2004). Affective and physiological responses to environmental noises and music. *International Journal of Psychophysiology*, *53*(2), 91–103.

Gorini, A., Griez, E., Petrova, A., & Riva, G. (2010). Assessment of the emotional responses produced by exposure to real food, virtual food and photographs of food in patients affected by eating disorders. *Annals of General Psychiatry*, *9*(1), 30.

Gorini, A., Mosso, J. L., Mosso, D., Pineda, E., Ruíz, N. L., Ramíez, M., ... Riva, G. (2009). Emotional response to virtual reality exposure across different cultures: the role of the attribution process. *Cyberpsychology & Behavior*, *12*(6), 699–705.

Götz, T., Bieg, M., & Hall, N. C. (2016). Assessing academic emotions via the experience sampling method. In *Methodological Advances in Research on Emotion and Education* (pp. 245–258). Springer.

Graves, R., Landis, T., & Goodglass, H. (1981). Laterality and sex differences for visual recognition of emotional and non-emotional words. *Neuropsychologia*, *19*(1), 95–102.

Griffith, D. M., Veech, J. A., & Marsh, C. J. (2016). Cooccur: probabilistic species co-occurrence analysis in R. *Journal of Statistical Software*, *69*(2), 1–17.

Griffiths, P. (2004). Towards a Machiavellian theory of emotional appraisal. In P. Cruse & D. Evans (Eds.), *Emotion, Evolution and Rationality* (pp. 89–105). Oxford: Oxford University Press.

Gross, J. J., & Levenson, R. W. (1995). Emotion elicitation using films. *Cognition & Emotion*, *9*(1), 87–108.

Gross, R. (2005). Face databases. In S. Li & A. Jain (Eds.), *Handbook of Face Recognition* (pp. 301–327). London: Springer Verlag.

Grühn, D., & Scheibe, S. (2008). Age-related differences in valence and arousal ratings of pictures from the International Affective Picture System (IAPS): Do ratings become more extreme with age? *Behavior Research Methods*, *40*(2), 512–521.

Grühn, D., & Sharifian, N. (2016). Lists of emotional stimuli. In H. L. Meiselman (Ed.), *Emotion Measurement* (pp. 145–164). Tokyo: Woodhead Publishing.

Hallahan, M., & Rosenthal, R. (1996). Statistical power: Concepts, procedures, and applications. *Behaviour Research and Therapy*, *34*(5), 489–499.

Hamann, S., & Mao, H. (2002). Positive and negative emotional verbal stimuli elicit activity in the left amygdala. *Neuroreport*, *13*(1), 15–19.

Hammond, K. R. (1948). Subject and object sampling—a note. *Psychological Bulletin*, *45*(6), 530.

Hammond, K. R. (1966). *The psychology of Egon Brunswik.* Oxford, England: Holt, Rinehart & Winston.

Hansen, B. B., & Klopfer, S. O. (2006). Optimal full matching and related designs via network flows. *Journal of Computational and Graphical Statistics*, *15*(3), 609–627.

Harmon-Jones, E., Amodio, D. M., & Zinner, L. R. (2007). Social psychological methods of emotion elicitation. In J. A. Coan & J. J. Allen (Eds.), *Handbook of Emotion Elicitation and Assessment.* (pp. 91–105). New York: Oxford university press.

Harmon-Jones, E., & Sigelman, J. (2001). State anger and prefrontal brain activity: evidence that insult-related relative left-prefrontal activation is associated with experienced anger and aggression. *Journal of Personality and Social Psychology*, *80*(5), 797.

Hartigan, J. A. (1975). *Clustering algorithms.* New York: Wiley.

Hartigan, J. A., & Wong, M. A. (1979). Algorithm AS 136: A k-means clustering algorithm. *Journal of Applied Statistics*, *28*(1), 100–108.

Harzing, A.-W. (2017). *Publish or Perish [Computer software].* Retrieved on December 22, 2016, from: http://www.harzing.com/resources/publish-or-perish.

Hautamäki, V., Cherednichenko, S., Kärkkäinen, I., Kinnunen, T., & Fränti, P. (2005). Improving k-means by outlier removal. In *Proceedings of the 14th Scandinavian Conference on Image Analysis* (pp. 978–987). Berlin: Springer-Verlag.

Hlavac, M. (2013). stargazer: LaTeX Code for Well-Formatted Regression and Summary Statistics Tables. *Harvard University, Cambridge, USA. R package version*, *3*(1).

Hofmann, W., & Patel, P. V. (2015). SurveySignal: A convenient solution for experience sampling research using participants' own smartphones. *Social Science Computer Review*, *33*(2), 235–253.

Hogarth, R. M. (2005). The challenge of representative design in psychology and economics. *Journal of Economic Methodology*, *12*(2), 253–263.

Hornik, K. (2005). A CLUE for CLUster ensembles. *Journal of Statistical Software*, *14*(12), 1–25.

Houtkamp, J. M. (2012). *Affective appraisal of virtual environments* (Unpublished doctoral dissertation). Utrecht University.

Howell, R. T., Rodzon, K. S., Kurai, M., & Sanchez, A. H. (2010). A validation of well-being and happiness surveys for administration via the Internet. *Behavior Research Methods*, *42*(3), 775–784.

Hubert, L., & Arabie, P. (1985). Comparing partitions. *Journal of Classification*, *2*(1), 193–218.

Hunter, E. M., Phillips, L. H., & MacPherson, S. E. (2010). Effects of age on cross-modal emotion perception. *Psychology and Aging*, *25*(4), 779.

Hupont, I., Gracia, J., Sanagustín, L., & Gracia, M. A. (2015). How do new visual immersive systems influence gaming QoE? A use case of serious gaming with Oculus Rift. In *The 7th International Workshop on Quality of Multimedia Experience (QoMEX)* (pp. 1–6).

Ijsselsteijn, W. A. (2002). Elements of a multi-level theory of presence: Phenomenology, mental processing and neural correlates. *Proceedings of PRESENCE*, *2002*, 245–259.

Irtel, H. (2008). *The PXLab Self-Assessment-Manikin Scales.* Website. (http://irtel.uni-mannheim.de/pxlab/demos/index_SAM.html. Retrieved at: March 13, 2017)

Isaacowitz, D. M., Löckenhoff, C. E., Lane, R. D., Wright, R., Sechrest, L., Riedel, R., & Costa, P. T. (2007). Age differences in recognition of emotion in lexical stimuli and facial expressions. *Psychology and Aging*, *22*(1), 147.

Isaacowitz, D. M., & Stanley, J. T. (2011). Bringing an ecological perspective to the study of aging and recognition of emotional facial expressions: Past, current, and future methods. *Journal of Nonverbal Behavior*, *35*(4), 261.

Ito, T. A., Cacioppo, J. T., & Lang, P. J. (1998). Eliciting affect using the International Affective Picture System: Trajectories through evaluative space. *Personality and Social Psychology Bulletin*, *24*(8), 855–879.

Izard, C. E. (1971). The face of emotion.

Jack, R. E., Blais, C., Scheepers, C., Schyns, P. G., & Caldara, R. (2009). Cultural confusions show that facial expressions are not universal. *Current Biology*, *19*(18), 1543–1548.

Jack, R. E., Garrod, O. G., Yu, H., Caldara, R., & Schyns, P. G. (2012). Facial expressions of emotion are not culturally universal. *Proceedings of the National Academy of Sciences*, *109*(19), 7241–7244.

Jain, A. K. (2010). Data clustering: 50 years beyond K-means. *Pattern Recognition Letters*, *31*(8), 651–666.

Jain, A. K., & Dubes, R. C. (1988). *Algorithms for clustering data.* Englewood Cliffs: Prentice Hall.

Jallais, C., & Gilet, A.-L. (2010). Inducing changes in arousal and valence: Comparison of two mood induction procedures. *Behavior Research Methods*, *42*(1), 318–325.

James, W. (1884). What is an emotion? *Mind*, *9*(34), 188–205.

Janschewitz, K. (2008). Taboo, emotionally valenced, and emotionally neutral word norms. *Behavior Research Methods*, *40*(4), 1065–1074.

Jennings, P. D., McGinnis, D., Lovejoy, S., & Stirling, J. (2000). Valence and arousal ratings for Velten mood induction statements. *Motivation and Emotion*, *24*(4), 285–297.

Jerit, J., Barabas, J., & Clifford, S. (2013). Comparing contemporaneous laboratory and field experiments on media effects. *Public Opinion Quarterly*, *77*(1), 256–282.

Kantardzic, M. (2011). *Data mining: concepts, models, methods, and algorithms (2nd ed.).* New Jersey: John Wiley & Sons.

Kappas, A., & Descôteaux, J. (2004). Les promesses et limites de l'étude de l'émotion en laboratoire. In G. Kirouac (Ed.), *Cognition et Émotions* (pp. 129–156). Coimbra: Imprensa da Universidade de Coimbra/Coimbra University Press.

Kaufman, L., & Rousseeuw, P. J. (2005). *Finding groups in data: an introduction to cluster analysis* (Vol. 344). New Jersey: John Wiley & Sons.

Keil, A., Bradley, M. M., Junghöfer, M., Russmann, T., Lowenthal, W., & Lang, P. J. (2007). Cross-modal attention capture by affective stimuli: evidence from event-related potentials. *Cognitive, Affective, & Behavioral Neuroscience*, *7*(1), 18–24.

Keil, A., & Miskovic, V. (2015). Human Emotions: A Conceptual Overview. In K. Babson & M. Feldner (Eds.), *Sleep and Affect: Assessment, Theory, and Clinical Implications* (pp. 23–44). Amsterdam: Academic Press.

Kendler, K. S., Myers, J., Prescott, C. A., & Neale, M. C. (2001). The genetic epidemiology of irrational fears and phobias in men. *Archives of General Psychiatry*, *58*(3), 257–265.

Kensinger, E. A. (2008). Age differences in memory for arousing and nonarousing emotional words. *The Journals of Gerontology Series B: Psychological Sciences and Social Sciences*, *63*(1), P13–P18.

Kensinger, E. A., & Corkin, S. (2004). Two routes to emotional memory: Distinct neural processes for valence and arousal. *Proceedings of the National Academy of Sciences*, *101*(9), 3310–3315.

Kiehl, K. A., Smith, A. M., Hare, R. D., Mendrek, A., Forster, B. B., Brink, J., & Liddle, P. F. (2001). Limbic abnormalities in affective processing by criminal psychopaths as revealed by functional magnetic resonance imaging. *Biological Psychiatry*, *50*(9), 677–684.

Kilteni, K., Normand, J.-M., Sanchez-Vives, M. V., & Slater, M. (2012). Extending body space in immersive virtual reality: a very long arm illusion. *PloS One*, *7*(7), e40867.

Kim, K., Rosenthal, M. Z., Zielinski, D. J., & Brady, R. (2014). Effects of virtual environment platforms on emotional responses. *Computer Methods and Programs in Biomedicine*, *113*(3), 882–893.

Kingstone, A., Smilek, D., & Eastwood, J. D. (2008). Cognitive ethology: A new approach for studying human cognition. *British Journal of Psychology*, *99*(3), 317–340.

Koenigsberg, H. W., Fan, J., Ochsner, K. N., Liu, X., Guise, K., Pizzarello, S., ... others (2010). Neural correlates of using distancing to regulate emotional responses to social situations. *Neuropsychologia*, *48*(6), 1813–1822.

Kooiman, C., Spinhoven, P., & Trijsburg, R. (2002). The assessment of alexithymia: a critical review of the literature and a psychometric study of the Toronto Alexithymia Scale-20. *Journal of Psychosomatic Research*, *53*(6), 1083–1090.

Kousta, S.-T., Vinson, D. P., & Vigliocco, G. (2009). Emotion words, regardless of polarity, have a processing advantage over neutral words. *Cognition*, *112*(3), 473–481.

Kroenke, K., Spitzer, R. L., & Williams, J. B. (2001). The PHQ-9. *Journal of General Internal Medicine*, *16*(9), 606–613.

Kroenke, K., Strine, T. W., Spitzer, R. L., Williams, J. B., Berry, J. T., & Mokdad, A. H. (2009). The PHQ-8 as a measure of current depression in the general population. *Journal of Affective Disorders*, *114*(1), 163–173.

Kuchinke, L., Võ, M. L.-H., Hofmann, M., & Jacobs, A. M. (2007). Pupillary responses during lexical decisions vary with word frequency but not emotional valence. *International Journal of Psychophysiology*, *65*(2), 132–140.

Kuliga, S. F., Thrash, T., Dalton, R. C., & Hölscher, C. (2015). Virtual reality as an empirical research tool—Exploring user experience in a real building and a corresponding virtual model. *Computers, Environment and Urban Systems*, *54*, 363–375.

Kunzmann, U., Kupperbusch, C. S., & Levenson, R. W. (2005). Behavioral inhibition and amplification during emotional arousal: a comparison of two age groups. *Psychology and Aging*, *20*(1), 144.

Kuppens, P. (2008). Individual differences in the relationship between pleasure and arousal. *Journal of Research in Personality*, *42*(4), 1053–1059.

Kuppens, P., Tuerlinckx, F., Russell, J. A., & Barrett, L. F. (2013). The relation between valence and arousal in subjective experience. *Psychological Bulletin*, *139*(4), 917.

Kuppens, P., Tuerlinckx, F., Yik, M., Koval, P., Coosemans, J., Zeng, K. J., & Russell, J. A. (2017). The relation between valence and arousal in subjective experience varies with personality and culture. *Journal of Personality*, *85*(4), 530–542.

Kuppens, P., Van Mechelen, I., Nezlek, J. B., Dossche, D., & Timmermans, T. (2007). Individual differences in core affect variability and their relationship to personality and psychological adjustment. *Emotion*, *7*(2), 262.

LaBar, K. S., & Phelps, E. A. (1998). Arousal-mediated memory consolidation: Role of the medial temporal lobe in humans. *Psychological Science*, *9*(6), 490–493.

Lane, R. D., Reiman, E. M., Bradley, M. M., Lang, P. J., Ahern, G. L., Davidson, R. J., & Schwartz, G. E. (1997). Neuroanatomical correlates of pleasant and unpleasant emotion. *Neuropsychologia*, *35*(11), 1437–1444.

Lang, P. J. (1980). Technology in mental health care delivery systems. In J. B. Sidowski, J. H. Johnson, & T. A. Williams (Eds.), *Behavioral Treatment and Bio-Behavioral Assessment: Computer Applications* (pp. 119–137). Norwood, NJ: Ablex.

Lang, P. J., Bradley, M. M., & Cuthbert, B. (2008). *International affective picture system (IAPS): Affective ratings of pictures and instruction manual.* (Tech. Rep.). University of Florida, Gainesville, FL.

Lang, P. J., Bradley, M. M., & Cuthbert, B. N. (1999). *International affective picture system (IAPS): Technical manual and affective ratings.* Gainesville, FL: The Center for Research in Psychophysiology, University of Florida.

Lange, C. G. (1885). The mechanism of the emotions. *The Classical Psychologists*, 672–684.

Larsen, J. T., Norris, C. J., & Cacioppo, J. T. (2003). Effects of positive and negative affect on electromyographic activity over zygomaticus major and corrugator supercilii. *Psychophysiology*, *40*(5), 776–785.

Larson, R., & Csikszentmihalyi, M. (2014). The experience sampling method. In M. Csikszentmihalyi (Ed.), *Flow and the Foundations of Positive Psychology* (pp. 21–34). Heidelberg: Springer.

Leary, M. R., & Kowalski, R. M. (1990). Impression management: A literature review and two-component model. *Psychological Bulletin*, *107*(1), 34.

LeDoux, J. (1998). *The emotional brain: The mysterious underpinnings of emotional life.* New York: Simon and Schuster.

Lench, H. C., Flores, S. A., & Bench, S. W. (2011). Discrete emotions predict changes in cognition, judgment, experience, behavior, and physiology: a meta-analysis of experimental emotion elicitations. *Psychological Bulletin*, *137*(5), 834.

Lepping, R. J., Atchley, R. A., & Savage, C. R. (2016). Development of a validated emotionally provocative musical stimulus set for research. *Psychology of Music*, *44*(5), 1012–1028.

Lessiter, J., Freeman, J., Keogh, E., & Davidoff, J. (2001). A cross-media presence questionnaire: The ITC-Sense of Presence Inventory. *Presence: Teleoperators and Virtual Environments*, *10*(3), 282–297.

Levenson, R. W., Ekman, P., & Friesen, W. V. (1990). Voluntary facial action generates emotion-specific autonomic nervous system activity. *Psychophysiology*, *27*(4), 363–384.

Lewis, P. A., Critchley, H., Rotshtein, P., & Dolan, R. (2007). Neural correlates of processing valence and arousal in affective words. *Cerebral Cortex*, *17*(3), 742–748.

Leys, C., Ley, C., Klein, O., Bernard, P., & Licata, L. (2013). Detecting outliers: do not use standard deviation around the mean, use absolute deviation around the median. *Journal of Experimental Social Psychology*, *49*(4), 764–766.

Lindquist, K. A., Gendron, M., Barrett, L. F., & Dickerson, B. C. (2014). Emotion perception, but not affect perception, is impaired with semantic memory loss. *Emotion*, *14*(2), 375.

Lithari, C., Frantzidis, C., Papadelis, C., Vivas, A. B., Klados, M., Kourtidou-Papadeli, C., ... Bamidis, P. (2010). Are females more responsive to emotional stimuli? A neurophysiological study across arousal and valence dimensions. *Brain Topography*, *23*(1), 27–40.

Lobbestael, J., Arntz, A., & Wiers, R. W. (2008). How to push someone's buttons: A comparison of four anger-induction methods. *Cognition & Emotion*, *22*(2), 353–373.

Loomis, J. M., Blascovich, J. J., & Beall, A. C. (1999). Immersive virtual environment technology as a basic research tool in psychology. *Behavior Research Methods, Instruments, & Computers*, *31*(4), 557–564.

Machajdik, J., & Hanbury, A. (2010). Affective image classification using features inspired by psychology and art theory. In *ACM Multimedia.*

MacQueen, J. (1967). Some methods for classification and analysis of multivariate observations. In L. Lecam & J. Neyman (Eds.), *Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability.* (Vol. 1(14), pp. 281–297). Berkeley, CA: The Regents of the University of California.

Marchewka, A., Żurawski, Ł., Jednoróg, K., & Grabowska, A. (2014). The Nencki Affective Picture System (NAPS): Introduction to a novel, standardized, wide-range, high-quality, realistic picture database. *Behavior Research Methods*, *46*(2), 596–610.

Martin, M. (1990). On the induction of mood. *Clinical Psychology Review*, *10*(6), 669–697. Retrieved from http://www.sciencedirect.com/science/article/pii/027273589090075L doi: 10.1016/0272-7358(90)90075-L

Mather, M. (2007). Emotional arousal and memory binding: An object-based framework. *Perspectives on Psychological Science*, *2*(1), 33–52.

Mather, M., & Nesmith, K. (2008). Arousal-enhanced location memory for pictures. *Journal of Memory and Language*, *58*(2), 449–464.

Mathôt, S., Schreij, D., & Theeuwes, J. (2012). OpenSesame: An open-source, graphical experiment builder for the social sciences. *Behavior Research Methods*, *44*(2), 314–324.

Mattek, A. M., Wolford, G. L., & Whalen, P. J. (2017). A mathematical model captures the structure of subjective affect. *Perspectives on Psychological Science*, *12*(3), 508–526.

Mauss, I. B., Levenson, R. W., McCarter, L., Wilhelm, F. H., & Gross, J. J. (2005). The tie that binds? Coherence among emotion experience, behavior, and physiology. *Emotion*, *5*(2), 175.

Mauss, I. B., & Robinson, M. D. (2009). Measures of emotion: A review. *Cognition and Emotion*, *23*(2), 209–237.

Mayer, J. D., Allen, I. P., & Beauregard, K. (1995). Mood inductions for four specific moods: A procedure employing guided imagery. *Journal of Mental imagery*, *19*(1-2), 133–150.

McCall, C., Hildebrandt, L. K., Hartmann, R., Baczkowski, B. M., & Singer, T. (2016). Introducing the Wunderkammer as a tool for emotion research: Unconstrained gaze and movement patterns in three emotionally evocative virtual worlds. *Computers in Human Behavior*, *59*, 93–107.

McMahan, A. (2003). Immersion, engagement and presence. *The Video Game Theory Reader*, *67*, 86.

Mehrabian, A. (1970). A semantic space for nonverbal behavior. *Journal of Consulting and Clinical Psychology*, *35*(2), 248.

Mehrabian, A. (1995). Framework for a comprehensive description and measurement of emotional states. *Genetic, Social, and General Psychology Monographs*.

Mehrabian, A. (1996). Pleasure-arousal-dominance: A general framework for describing and measuring individual differences in temperament. *Current Psychology*, *14*(4), 261–292.

Mehrabian, A., & Russell, J. A. (1974). *An approach to environmental psychology.* Cambrige, MA: MIT Press.

Meilă, M. (2007). Comparing clusterings—an information based distance. *Journal of Multivariate Analysis*, *98*(5), 873–895.

Mikels, J. A., Fredrickson, B. L., Larkin, G. R., Lindberg, C. M., Maglio, S. J., & Reuter-Lorenz, P. A. (2005). Emotional category data on images from the International Affective Picture System. *Behavior Research Methods*, *37*(4), 626–630.

Mikels, J. A., Larkin, G. R., Reuter-Lorenz, P. A., & Carstensen, L. L. (2005). Divergent trajectories in the aging mind: Changes in working memory for affective versus visual information with age. *Psychology and Aging*, *20*(4), 542.

Miles, L., & Johnston, L. (2007). Detecting happiness: Perceiver sensitivity to enjoyment and non-enjoyment smiles. *Journal of Nonverbal Behavior*, *31*(4), 259–275.

Miller, M. W., Patrick, C. J., & Levenston, G. K. (2002). Affective imagery and the startle response: Probing mechanisms of modulation during pleasant scenes, personal experiences, and discrete negative emotions. *Psychophysiology*, *39*(4), 519–529.

Mitchell, G. (2012). Revisiting Truth or Triviality The External Validity of Research in the Psychological Laboratory. *Perspectives on Psychological Science*, *7*(2), 109–117.

Mooi, E., & Sarstedt, M. (2011). *A concise guide to market research: The process, data, and methods using IBM SPSS statistics.* Berlin: Springer.

Moors, A. (2009). Theories of emotion causation: A review. *Cognition and Emotion*, *23*(4), 625–662.

Moors, A., Ellsworth, P. C., Scherer, K. R., & Frijda, N. H. (2013). Appraisal theories of emotion: State of the art and future development. *Emotion Review*, *5*(2), 119–124.

Most, S. B., Chun, M. M., Widders, D. M., & Zald, D. H. (2005). Attentional rubbernecking: Cognitive control and personality in emotion-induced blindness. *Psychonomic Bulletin & Review*, *12*(4), 654–661.

Mun, E. Y., von Eye, A., Bates, M. E., & Vaschillo, E. G. (2008). Finding groups using model-based cluster analysis: heterogeneous emotional self-regulatory processes and heavy alcohol use risk. *Developmental Psychology*, *44*(2), 481.

Nelson, N. L., & Russell, J. A. (2016). A facial expression of pax: Assessing children's "recognition" of emotion from faces. *Journal of Experimental Child Psychology*, *141*, 49–64.

Neumann, R., & Strack, F. (2000). " Mood contagion": the automatic transfer of mood between persons. *Journal of Personality and Social Psychology*, *79*(2), 211.

Niedenthal, P. M., Halberstadt, J. B., & Innes-Ker, Å. H. (1999). Emotional response categorization. *Psychological Review*, *106*(2), 337.

Nielsen, L., & Kaszniak, A. (2007). Conceptual, theortical, and methodological issues in inferring subjective emotion experience. In J. A. Coan & J. J. Allen (Eds.), *Handbook of Emotion Elicitation and Assessment.* (pp. 361–375). New York: Oxford university press.

Norris, C. J., Larsen, J. T., Crawford, L. E., & Cacioppo, J. T. (2011). Better (or worse) for some than others: Individual differences in the positivity offset and negativity bias. *Journal of Research in Personality*, *45*(1), 100–111.

Noulhiane, M., Mella, N., Samson, S., Ragot, R., & Pouthas, V. (2007). How emotional auditory stimuli modulate time perception. *Emotion*, *7*(4), 697.

Nummenmaa, L., & Niemi, P. (2004). Inducing affective states with success-failure manipulations: a meta-analysis. *Emotion*, *4*(2), 207–214.

Ortony, A., Clore, G. L., & Collins, A. (1990). *The cognitive structure of emotions.* Cambridge: Cambridge University Press.

Ortony, A., & Turner, T. J. (1990). What's basic about basic emotions? *Psychological Review*, *97*(3), 315.

Osgood, C. E. (1952). The nature and measurement of meaning. *Psychological Bulletin*, *49*(3), 197.

Osgood, C. E., Suci, G. J., & Tannenbaum, P. H. (1964). *The measurement of meaning.* University of Illinois Press.

Pallavicini, F., Cipresso, P., Raspelli, S., Grassi, A., Serino, S., Vigna, C., ... Riva, G. (2013). Is virtual reality always an effective stressors for exposure treatments? Some insights from a controlled trial. *BMC Psychiatry*, *13*(1), 52.

Parker, J. D., Taylor, G. J., & Bagby, R. M. (2003). The 20-Item Toronto Alexithymia Scale: III. Reliability and factorial validity in a community population. *Journal of Psychosomatic Research*, *55*(3), 269–275.

Parsons, T. D. (2015). Virtual reality for enhanced ecological validity and experimental control in the clinical, affective and social neurosciences. *Frontiers in Human Neuroscience*, *9*, 660.

Parsons, T. D., Gaggioli, A., & Riva, G. (2017). Virtual reality for research in social neuroscience. *Brain Sciences*, *7*(4), 42.

Parsons, T. D., & Rizzo, A. A. (2008). Affective outcomes of virtual reality exposure therapy for anxiety and specific phobias: A meta-analysis. *Journal of Behavior Therapy and Experimental Psychiatry*, *39*(3), 250–261.

Partala, T. (2011). Psychological needs and virtual worlds: Case Second Life. *International Journal of Human-Computer Studies*, *69*(12), 787–800.

Partala, T., & Surakka, V. (2003). Pupil size variation as an indication of affective processing. *International Journal of Human-Computer Studies*, *59*(1), 185–198.

Perri, R. L., Berchicci, M., Lucci, G., Cimmino, R. L., Bello, A., & Di Russo, F. (2014). Getting ready for an emotion: specific premotor brain activities for self-administered emotional pictures. *Frontiers in Behavioral Neuroscience*, *8*.

Plutchik, R. (1980). A general psychoevolutionary theory of emotion. In R. Plutchik & H. Kellerman (Eds.), *Emotion: Theory, Research, and Experience* (Vol. 1, pp. 3–31). New York: Academic Press.

Plutchik, R. (2009). Emotions: A general psychoevolutionary theory. In K. Scherer & P. Ekman (Eds.), *Approaches to Emotion* (pp. 197–219). New York: Psychology press taylor & Francis Group.

Posner, J., Russell, J. A., & Peterson, B. S. (2005). The circumplex model of affect: An integrative approach to affective neuroscience, cognitive development, and psychopathology. *Development and Psychopathology*, *17*(03), 715–734.

Prat, M. (2006). *Processus cognitifs et émotions.* Paris: Editions L'Harmattan.

R Core Team. (2015). *R: A language and environment for statistical computing (Version 3.2.1) [Computer software].* (Tech. Rep.). R Core Team. (Retrieved from: http://www.R-project.org/.)

Redondo, J., Fraga, I., Padrón, I., & Comesaña, M. (2007). The Spanish adaptation of ANEW (affective norms for English words). *Behavior Research Methods*, *39*(3), 600–605.

Redondo, J., Fraga, I., Padrón, I., & Piñeiro, A. (2008). Affective ratings of sound stimuli. *Behavior Research Methods*, *40*(3), 784–790.

Richardson, A., & Taylor, C. C. (1982). Vividness of memory imagery and self-induced mood change. *British Journal of Clinical Psychology*, *21*(2), 111–117.

Riegel, M., Wierzba, M., Wypych, M., Żurawski, Ł., Jednoróg, K., Grabowska, A., & Marchewka, A. (2015). Nencki affective word list (NAWL): The cultural adaptation of the Berlin affective word list–reloaded (BAWL-R) for Polish. *Behavior Research Methods*, *47*(4), 1222–1236.

Riva, G., Mantovani, F., Capideville, C. S., Preziosa, A., Morganti, F., Villani, D., . . . Alcañiz, M. (2007). Affective interactions using virtual reality: the link between presence and emotions. *CyberPsychology & Behavior*, *10*(1), 45–56.

Roberts, N. A., & Levenson, R. W. (2006). Subjective, behavioral, and physiological reactivity to ethnically matched and ethnically mismatched film clips. *Emotion*, *6*(4), 635.

Rosch, E., Mervis, C. B., Gray, W. D., Johnson, D. M., & Boyes-Braem, P. (1976). Basic objects in natural categories. *Cognitive Psychology*, *8*(3), 382–439.

Rouby, C., Fournel, A., & Bensafi, M. (2016). The Role of the Senses in Emotion. In H. L. Meiselman (Ed.), *Emotion Measurement* (pp. 65–81). Tokyo: Woodhead Publishing.

Ruef, A., & Levenson, R. (2007). Continuous measurement of emotion. The affect rating dial. In J. A. Coan & J. J. Allen (Eds.), *Handbook of Emotion Elicitation and Assessment.* (pp. 287–297). New York: Oxford university press.

Russell, J. A. (1978). Evidence of convergent validity on the dimensions of affect. *Journal of Personality and Social Psychology*, *36*(10), 1152.

Russell, J. A. (1980). A circumplex model of affect. *Journal of Personality and Social Psychology*, *39*, 1161–1178.

Russell, J. A. (1991). In defense of a prototype approach to emotion concepts. *Journal of Personality and Social Psychology*, *60*(1), 37.

Russell, J. A. (2003). Core affect and the psychological construction of emotion. *Psychological Review*, *110*(1), 145.

Russell, J. A. (2005). Emotion in human consciousness is built on core affect. *Journal of Consciousness Studies*, *12*(8-9), 26–42.

Russell, J. A., & Barrett, L. F. (1999). Core affect, prototypical emotional episodes, and other things called emotion: dissecting the elephant. *Journal of Personality and Social Psychology*, *76*(5), 805.

Russell, J. A., & Mehrabian, A. (1977). Evidence for a three-factor theory of emotions. *Journal of Research in Personality*, *11*(3), 273–294.

Russell, J. A., Weiss, A., & Mendelsohn, G. A. (1989). Affect grid: A single-item scale of pleasure and arousal. *Journal of Personality and Social Psychology*, *57*(3), 493–502.

Salas, C. E., Radovic, D., & Turnbull, O. H. (2012). Inside-out: comparing internally generated and externally generated basic emotions. *Emotion*, *12*(3), 568.

Schachter, S., & Singer, J. (1962). Cognitive, social, and physiological determinants of emotional state. *Psychological Review*, *69*(5), 379.

Schaefer, A., Nils, F., Sanchez, X., & Philippot, P. (2010). Assessing the effectiveness of a large database of emotion-eliciting films: A new tool for emotion researchers. *Cognition and Emotion*, *24*(7), 1153–1172.

Schaefer, H. S., Larson, C. L., Davidson, R. J., & Coan, J. A. (2014). Brain, body, and cognition: Neural, physiological and self-report correlates of phobic and normative fear. *Biological Psychology*, *98*, 59–69.

Scherer, K. R. (2001). Appraisal considered as a process of multilevel sequential checking. In K. R. Scherer, A. Schorr, & T. Johnstone (Eds.), *Appraisal Processes in Emotion* (pp. 92–120). New York: Oxford University Press.

Scherer, K. R., & Ekman, P. (2009). On the nature and function of emotion: A component process approach. In *Approaches to Emotion* (pp. 293–317). New York: Psychology Press, Taylor & Francis Group.

Schmid, H. (2013). Probabilistic part-ofispeech tagging using decision trees. In *New Methods in Language Processing* (p. 154).

Schubert, T., Friedmann, F., & Regenbrecht, H. (2001). The experience of presence: Factor analytic insights. *Presence*, *10*(3), 266–281.

Scollon, C. N., Kim-Prieto, C., & Diener, E. (2003). Experience sampling: Promises and pitfalls, strengths and weaknesses. *Journal of Happiness Studies*, *4*(1), 5–34.

Scott, G. G., O'Donnell, P. J., Leuthold, H., & Sereno, S. C. (2009). Early emotion word processing: Evidence from event-related potentials. *Biological Psychology*, *80*(1), 95–104.

Serrano, B., Botella, C., Baños, R. M., & Alcañiz, M. (2013). Using virtual reality and mood-induction procedures to test products with consumers of ceramic tiles. *Computers in Human Behavior*, *29*(3), 648–653.

Shaver, P., Schwartz, J., Kirson, D., & O'Connor, C. (1987). Emotion knowledge: Further exploration of a prototype approach. *Journal of Personality and Social Psychology*, *52*(6), 1061.

Slater, M. (1999). Measuring presence: A response to the Witmer and Singer presence questionnaire. *Presence*, *8*(5), 560–565.

Slater, M., Spanlang, B., Sanchez-Vives, M. V., & Blanke, O. (2010). First person experience of body transfer in virtual reality. *PloS One*, *5*(5), e10564.

Slater, M., & Wilbur, S. (1997). A framework for immersive virtual environments (FIVE): Speculations on the role of presence in virtual environments. *Presence: Teleoperators & Virtual Environments*, *6*(6), 603–616.

Sloman, A. (2004). What are emotion theories about. In *Architectures for Modeling Emotion: Cross-Disciplinary Foundations. AAAI Spring Symposium Technical Report* (pp. 128–134).

Soares, A. P., Comesaña, M., Pinheiro, A. P., Simões, A., & Frade, C. S. (2012). The adaptation of the Affective Norms for English words (ANEW) for European Portuguese. *Behavior Research Methods*, *44*(1), 256–269.

Solomon, R. (2008). The Philosophy of Emotions. In M. Lewis, J. M. Haviland-Jones, & L. F. Barrett (Eds.), *Handbook of Emotions* (pp. 3–17). New York: The Guilford Press.

Sorabji, R. (2000). *Emotion and peace of mind: From stoic agitation to Christian temptation.* Oxford: Oxford University Press.

Soussignan, R. (2002). Duchenne smile, emotional experience, and autonomic reactivity: a test of the facial feedback hypothesis. *Emotion*, *2*(1), 52.

Stadthagen-González, H., Ferré, P., Pérez-Sánchez, M. A., Imbault, C., & Hinojosa, J. A. (2017). Norms for 10,491 Spanish words for five discrete emotions: Happiness, disgust, anger, fear, and sadness. *Behavior Research Methods*, 1–10.

Stadthagen-Gonzalez, H., Imbault, C., Pérez Sánchez, M. A., & Brysbaert, M. (2017). Norms of valence and arousal for 14,031 Spanish words. *Behavior Research Methods*, *49*(1), 111–123. Retrieved from http://dx.doi.org/10.3758/s13428-015-0700-2 doi: 10.3758/s13428-015-0700-2

Stein, C. (2016). Virtual Reality Design: How Head-Mounted Displays Change Design Paradigms of Virtual Reality Worlds. *MediaTropes*, *6*(1), 52–85.

Stemmler, G. (2003). Methodological considerations in the psychophysiological study of emotion. In R. Davidson, K. R. Scherer, & H. H. Goldsmith (Eds.), *Handbook of Affective Sciences* (pp. 225–255). Oxford: Oxford University Press.

Stevenson, R. A., & James, T. W. (2008). Affective auditory stimuli: Characterization of the International Affective Digitized Sounds (IADS) by discrete emotional categories. *Behavior Research Methods*, *40*(1), 315–321.

Stevenson, R. A., Mikels, J. A., & James, T. W. (2007). Characterization of the affective norms for English words by discrete emotional categories. *Behavior Research Methods*, *39*(4), 1020–1024.

Stins, J. F., & Beek, P. J. (2007). Effects of affective picture viewing on postural control. *BMC Neuroscience*, *8*(1), 83.

Strauss, G. P., & Allen, D. N. (2008). Emotional intensity and categorisation ratings for emotional and nonemotional words. *Cognition and Emotion*, *22*(1), 114–133.

Streiner, D. L., & Norman, G. R. (2011). Correction for multiple testing: is there a resolution? *Chest*, *140*(1), 16–18.

Stuart, K. (2016). *Why do we love Nathan Drake? Uncharted 4 designer explains all.* Website. (https://www.theguardian.com/technology/2016/apr/07/nathan-drake-uncharted-4-naughty-dog. Retrieved at: October 30, 2017)

Suk, H.-J. (2006). *Color and emotion: A study on the affective judgment of color across media and in relation to visual stimuli* (Unpublished doctoral dissertation). Social Science.

Suk, H.-J., & Irtel, H. (2010). Emotional response to color across media. *Color Research & Application*, *35*(1), 64–77.

Sutherland, G., Newman, B., & Rachman, S. (1982). Experimental investigations of the relations between mood and intrusive unwanted cognitions. *British Journal of Medical Psychology*, *55*(2), 127–138.

Tan, C. T., Leong, T. W., Shen, S., Dubravs, C., & Si, C. (2015). Exploring gameplay experiences on the Oculus Rift. In *Proceedings of the 2015 Annual Symposium on Computer-Human Interaction in Play* (pp. 253–263).

Tannenbaum, P. H., & Zillmann, D. (1975). Emotional arousal in the facilitation of aggression through communication. *Advances in Experimental Social Psychology*, *8*, 149–192.

Taylor, G. J., Bagby, R. M., & Parker, J. D. (2003). The 20-Item Toronto Alexithymia Scale: IV. Reliability and factorial validity in different languages and cultures. *Journal of Psychosomatic Research*, *55*(3), 277–283.

Thai, S., & Page-Gould, E. (2014). *ExperienceSampler: An open-source experience sampling smartphone app.* Website. (http://www.experiencesampler.com/. Retrieved at: December 06, 2016)

Thai, S., & Page-Gould, E. (2017). ExperienceSampler: An Open-Source Scaffold for Building Smartphone Apps for Experience Sampling. *Psychological Methods*. Retrieved from http://dx.doi.org/10.1037/met0000151 doi: 10.1037/met0000151

Thierry, G., & Roberts, M. V. (2007). Event-related potential study of attention capture by affective sounds. *Neuroreport*, *18*(3), 245–248.

Thompson, E. R. (2007). Development and validation of an internationally reliable short-form of the positive and negative affect schedule (PANAS). *Journal of Cross-Cultural Psychology*, *38*(2), 227–242.

Tiedens, L. Z., & Linton, S. (2001). Judgment under emotional certainty and uncertainty: the effects of specific emotions on information processing. *Journal of Personality and Social Psychology*, *81*(6), 973.

Tolman, E. C., & Brunswik, E. (1935). The organism and the causal texture of the environment. *Psychological Review*, *42*(1), 43.

Tomaszczyk, J. C., Fernandes, M. A., & MacLeod, C. M. (2008). Personal relevance modulates the positivity bias in recall of emotional pictures in older adults. *Psychonomic Bulletin & Review*, *15*(1), 191–196.

Tomkins, S. S. (1980). Affect as amplification: Some modifications in theory. In R. Plutchik & H. Kellerman (Eds.), *Emotion: Theory, Research, and Experience* (Vol. 1, pp. 141–164). New York: Academic Press.

Tremblay, A., & Ransijn, J. (2015). *Package 'LMERConvenienceFunctions'* (Tech. Rep.). Dalhousie University.

Trull, T. J., & Ebner-Priemer, U. (2013). Ambulatory assessment. *Annual Review of Clinical Psychology*, *9*, 151–176.

Uhrig, M. K., Trautmann, N., Baumgärtner, U., Treede, R.-D., Henrich, F., Hiller, W., & Marschall, S. (2016). Emotion elicitation: a comparison of pictures and films. *Frontiers in Psychology*, *7*.

Vaish, A., Grossmann, T., & Woodward, A. (2008). Not all emotions are created equal: the negativity bias in social-emotional development. *Psychological Bulletin*, *134*(3), 383.

Velten, E. (1968). A laboratory task for induction of mood states. *Behaviour Research and Therapy*, *6*(4), 473–482.

Verona, E., Patrick, C. J., Curtin, J. J., Bradley, M. M., & Lang, P. J. (2004). Psychopathy and physiological response to emotionally evocative sounds. *Journal of Abnormal Psychology*, *113*(1), 99.

Võ, M. L. H., Conrad, M., Kuchinke, L., Urton, K., Hofmann, M. J., & Jacobs, A. M. (2009). The Berlin Affective Word List Reloaded (BAWL-R). *Behavior Research Methods*, *41*(2), 534–538. Retrieved from http://dx.doi.org/10.3758/BRM.41.2.534 doi: 10.3758/BRM.41.2.534

Wang, W., & He, Q. (2008). A survey on emotional semantic image retrieval. In *The 15th IEEE International Conference on Image Processing* (pp. 117–120).

Warriner, A. B., Kuperman, V., & Brysbaert, M. (2013). Norms of valence, arousal, and dominance for 13,915 English lemmas. *Behavior Research Methods*, *45*(4), 1191–1207.

Watson, D., & Clark, L. A. (1999). The PANAS-X: Manual for the positive and negative affect schedule-expanded form [Computer software manual].

Watson, D., Clark, L. A., & Tellegen, A. (1988). Development and validation of brief measures of positive and negative affect: the PANAS scales. *Journal of Personality and Social Psychology*, *54*(6), 1063.

Wechsler, D. (1997). WAIS-III, Weschler adult intelligence scale - third edition. Administration and scoring manual [Computer software manual]. San Antonio, TX: The Psychological Corporation.

Weigold, A., Weigold, I. K., & Russell, E. J. (2013). Examination of the equivalence of self-report survey-based paper-and-pencil and internet data collection methods. *Psychological Methods*, *18*(1), 53.

Westermann, R., Spies, K., Stahl, G., & Hesse, F. W. (1996). Relative effectiveness and validity of mood induction procedures: A meta-analysis. *European Journal of Social Psychology*, *26*(4), 557–580.

Widen, S. C. (2013). Children's interpretation of facial expressions: The long path from valence-based to specific discrete categories. *Emotion Review*, *5*(1), 72–77.

Widen, S. C., & Russell, J. A. (2015). Do Dynamic Facial Expressions Convey Emotions to Children Better Than Do Static Ones? *Journal of Cognition and Development*, *16*(5), 802–811.

Wilhelm, F. H., & Grossman, P. (2010). Emotions beyond the laboratory: Theoretical fundaments, study design, and analytic strategies for advanced ambulatory assessment. *Biological Psychology*, *84*(3), 552–569.

Wilhelm, F. H., Roth, W. T., & Sackner, M. A. (2003). The LifeShirt: an advanced system for ambulatory measurement of respiratory and cardiac function. *Behavior Modification*, *27*(5), 671–691.

Williamson, S., Harpur, T. J., & Hare, R. D. (1991). Abnormal processing of affective words by psychopaths. *Psychophysiology*, *28*(3), 260–273.

Wolak, M. E., Fairbairn, D. J., & Paulsen, Y. R. (2012). Guidelines for Estimating Repeatability. *Methods in Ecology and Evolution 3(1):129-137*.

Wu, J. (2012). *Advances in K-means clustering: a data mining thinking.* Berlin: Springer Verlag.

Wundt, W. M. (1896). *Grundriss der psychologie.* Leibzig: Engelmann.

Xing, C., & Isaacowitz, D. M. (2006). Aiming at happiness: How motivation affects attention to and memory for emotional images. *Motivation and Emotion*, *30*(3), 243–250.

Xu, R., & Wunsch, D. (2009). *Clustering* (Vol. 10). New Jersey: John Wiley & Sons.

Yik, M. S., Russell, J. A., & Barrett, L. F. (1999). Structure of self-reported current affect: Integration and beyond. *Journal of Personality and Social Psychology*, *77*(3), 600.

Young, G. (2010). Virtually real emotions and the paradox of fiction: Implications for the use of virtual environments in psychological research. *Philosophical Psychology*, *23*(1), 1–21.

Young, I. T. (1977). Proof without prejudice: use of the Kolmogorov-Smirnov test for the analysis of histograms from flow systems and other sources. *Journal of Histochemistry & Cytochemistry*, *25*(7), 935–941.

Zajonc, R. B. (1968). Attitudinal effects of mere exposure. *Journal of Personality and Social Psychology*, *9*(2p2), 1.

Zajonc, R. B. (2001). Mere exposure: A gateway to the subliminal. *Current Directions in Psychological Science*, *10*(6), 224–228.

Zelenski, J. M., & Larsen, R. J. (2000). The distribution of basic emotions in everyday life: A state and trait perspective from experience sampling data. *Journal of Research in Personality*, *34*(2), 178–197.

Zentner, M., & Eerola, T. (2010). Self-report measures and models. In P. N. Juslin & J. Sloboda (Eds.), *Handbook of Music and Emotion: Theory, Research, Applications* (pp. 187–222). Oxford: Oxford University Press.

Zhang, X., Hui, W. Y., & Barrett, L. F. (2014). How does this make you feel? A comparison of four affect induction procedures. *Frontiers in Psychology*, *5*.

Zhao, X., & Sandelin, A. (2012). GMD: measuring the distance between histograms with applications on high-throughput sequencing reads. *Bioinformatics*, *28*(8), 1164–1165.

Zillmann, D., Katcher, A. H., & Milavsky, B. (1972). Excitation transfer from physical exercise to subsequent aggressive behavior. *Journal of Experimental Social Psychology*, *8*(3), 247–259.

Part VIII

Appendices

Appendix A

Clustered set of IAPS images

No	ImageCode	Description	Valence	Arousal	Dominance1	Cluster	Uncertainty	Under75thQuantile
1	1930.0	Shark	3.79	6.42	3.19	1	0.000000002	TRUE
2	1050.0	Snake	3.46	6.87	3.08	1	0.00000002	TRUE
3	5972.0	Tornado	3.85	6.34	3.49	1	0.000000054	TRUE
4	1931.0	Shark	4	6.8	3.51	1	0.000000127	TRUE
5	1321.0	Bear	4.32	6.64	3.51	1	0.000000522	TRUE
6	1070.0	Snake	3.96	6.16	3.71	1	0.000000535	TRUE
7	5971.0	Tornado	3.49	6.65	3.3	1	0.000000594	TRUE
8	1033.0	Snake	3.87	6.13	3.73	1	0.000000788	TRUE
9	9831.0	Cigarette	2.95	4.61	6.04	1	0.000000861	TRUE
10	9832.0	Cigarettes	2.94	4.46	5.53	1	0.000001447	TRUE
11	9230.0	OilFire	3.89	5.77	3.73	1	0.00000145	TRUE
12	1300.0	PitBull	3.55	6.79	3.49	1	0.000001618	TRUE
13	1052.0	Snake	3.5	6.52	3.36	1	0.000001646	TRUE
14	5940.0	Lava	4.23	6.29	3.7	1	0.000001774	TRUE
15	1040.0	Snake	3.99	6.25	3.92	1	0.000003778	TRUE
16	1120.0	Snake	3.79	6.93	3.87	1	0.000004027	TRUE
17	1110.0	Snake	3.84	5.96	3.84	1	0.000004479	TRUE
18	1051.0	Snake	3.8	5.95	3.8	1	0.000005692	TRUE
19	1113.0	Snake	3.81	6.06	3.91	1	0.000010568	TRUE
20	1022.0	Snake	4.26	6.02	3.9	1	0.000010914	TRUE
21	1304.0	AttackDog	3.37	6.37	3.29	1	0.000012124	TRUE
22	1302.0	Dog	4.21	6	4.04	1	0.000025488	TRUE
23	1101.0	Snake	4.1	5.83	4.13	1	0.000026997	TRUE
24	6832.0	Police	4.02	5.51	4.16	1	0.000029099	TRUE
25	1090.0	Snake	3.7	5.88	3.82	1	0.000039081	TRUE
26	1019.0	Snake	3.95	5.77	4.23	1	0.000045065	TRUE
27	3022.0	Scream	3.7	5.88	3.84	1	0.00004594	TRUE
28	2681.0	Police	4.04	4.97	3.84	1	0.000066458	TRUE
29	6000.0	Prison	4.04	4.91	3.77	1	0.000083389	TRUE
30	1200.0	Spider	3.95	6.03	4.33	1	0.000091131	TRUE
31	8480.0	BikerOnFire	3.7	6.28	4	1	0.000091638	TRUE
32	9008.0	Needle	3.47	4.45	5.3	1	0.000124739	TRUE
33	9403.0	Soldiers	3.51	5.62	3.49	1	0.000129083	TRUE
34	9395.0	Dishes	3.21	4.22	5.08	1	0.000135734	TRUE
35	1301.0	Dog	3.7	5.77	3.96	1	0.000141499	TRUE
36	9582.0	DentalExam	4.18	5.29	4.33	1	0.000175678	TRUE
37	2900.1	CryingBoy	2.56	4.61	4.83	1	0.000193844	TRUE
38	3211.0	Surgery	4.15	5.72	4.4	1	0.000209757	TRUE
39	1080.0	Snake	4.24	5.69	4.33	1	0.000210358	TRUE
40	1201.0	Spider	3.55	6.36	3.87	1	0.00024905	TRUE
41	2661.0	Baby	3.9	5.76	4.48	1	0.000291946	TRUE
42	3360.0	Fetus	3.78	5.39	4.18	1	0.000322038	TRUE
43	7640.0	Skyscraper	5	6.03	3.82	1	0.000328586	TRUE
44	9101.0	Cocaine	3.62	4.02	5.35	1	0.000341873	TRUE
45	6190.0	Aimedgun	3.57	5.64	3.77	1	0.000344687	TRUE
46	9290.0	Garbage	2.88	4.4	4.9	1	0.000349866	TRUE
47	2722.0	Jail	3.47	3.52	5.34	1	0.00035068	TRUE
48	3280.0	DentalExam	3.72	5.39	4.06	1	0.000402822	TRUE
49	9390.0	Dishes	3.67	4.14	5.22	1	0.000424199	TRUE
50	3310.0	Incubator	4.37	5.43	4.32	1	0.000450603	TRUE
51	1270.0	Roach	3.68	4.77	5.25	1	0.000465498	TRUE
52	9270.0	ToxicWaste	3.72	5.24	4.04	1	0.000512942	TRUE
53	9930.0	ShipWave	3.12	5.71	2.97	1	0.000550431	TRUE
54	6211.0	Attack	3.62	5.9	4.03	1	0.000552891	TRUE
55	1275.0	Roaches	3.3	4.81	5.11	1	0.000697465	TRUE
56	2115.0	PiercedFace	3.83	4.98	4.87	1	0.000842234	TRUE

No	ImageCode	Description	Valence	Arousal	Dominance1	Cluster	Uncertainty	Under75thQuantile
57	7023.0	Garbage	3.8	4.17	5.16	1	0.000868614	TRUE
58	6010.0	Jail	3.73	3.95	5.08	1	0.000886771	TRUE
59	9331.0	HomelessMan	2.87	3.85	4.72	1	0.000908444	TRUE
60	1505.0	DogRace	4.13	4.73	4.49	1	0.000975163	TRUE
61	1280.0	Rat	3.66	4.93	5.05	1	0.000982533	TRUE
62	3210.0	Surgery	4.49	5.39	4.3	1	0.000999873	TRUE
63	7360.0	FliesOnPie	3.59	5.11	5.21	1	0.001034584	TRUE
64	9445.0	Skeleton	3.87	4.49	4.51	1	0.001040775	TRUE
65	9594.0	Injection	3.76	5.17	4.43	1	0.00107692	TRUE
66	1030.0	Snake	4.3	5.46	4.56	1	0.001082138	TRUE
67	7521.0	Hospital	3.92	4.38	4.85	1	0.001090492	TRUE
68	9630.0	Bomb	2.96	6.06	2.98	1	0.001128473	TRUE
69	3250.0	OpenChest	3.78	6.29	4.45	1	0.001137449	TRUE
70	7520.0	Hospital	3.83	4.57	4.42	1	0.001200123	TRUE
71	9045.0	NativeFem	3.75	3.89	5.03	1	0.001259169	TRUE
72	6314.0	Attack	4.09	4.6	4.58	1	0.001314456	TRUE
73	8231.0	Boxer	3.77	5.24	4.68	1	0.001333691	TRUE
74	6240.0	Gun	3.79	5.27	4.9	1	0.001354821	TRUE
75	5920.0	Volcano	5.16	6.23	3.95	1	0.001373347	TRUE
76	6550.0	Attack	2.73	7.09	3.01	1	0.001374375	TRUE
77	9110.0	Puddle	3.76	3.98	4.88	1	0.001418885	TRUE
78	4635.0	Prostitute	3.86	4.23	4.7	1	0.001501331	TRUE
79	9490.0	Corpse	3.6	5.57	4.12	1	0.001687228	TRUE
80	9005.0	HIVTattoo	3.69	5.18	4.3	1	0.001696588	TRUE
81	2457.0	CryingBoy	3.2	4.94	5.02	1	0.001832471	TRUE
82	2110.0	AngryFace	3.71	4.53	4.66	1	0.001846255	TRUE
83	9265.0	HungMan	2.6	4.34	4.6	1	0.00184999	TRUE
84	6930.0	Missiles	4.39	4.88	4	1	0.001920927	TRUE
85	7079.0	Waste	3.81	4.47	4.19	1	0.001956767	TRUE
86	7078.0	Bucket	3.79	3.69	5.41	1	0.001957495	TRUE
87	6610.0	Gun	3.6	5.06	4.97	1	0.001961732	TRUE
88	9440.0	Skulls	3.67	4.55	4.69	1	0.002060538	TRUE
89	9291.0	Garbage	2.93	4.38	4.75	1	0.002078702	TRUE
90	9182.0	Horses	3.52	4.98	4.97	1	0.002108092	TRUE
91	2312.0	Mother	3.71	4.02	4.72	1	0.002218706	TRUE
92	3190.0	Scar	3.69	5.01	4.53	1	0.002287322	TRUE
93	1310.0	Leopard	4.6	6	4.37	1	0.002430068	TRUE
94	9404.0	Soldiers	3.71	4.67	4.48	1	0.002490603	TRUE
95	7092.0	Scale	4.05	4.38	4.99	1	0.002618894	TRUE
96	9596.0	Injection	3.65	5.13	4.31	1	0.00268586	TRUE
97	2130.0	Woman	4.08	5.02	5.1	1	0.002801185	TRUE
98	7137.0	CarDamage	4.3	4.81	4.5	1	0.002962109	TRUE
99	2039.0	Woman	3.65	3.46	5.06	1	0.003079544	TRUE
100	9080.0	Wires	4.07	4.36	4.05	1	0.003146668	TRUE
101	9150.0	Matador	4.54	5.31	4.48	1	0.003297683	TRUE
102	9090.0	Exhaust	3.625	4.385	4.615	1	0.003561929	TRUE
103	9623.0	Fire	3.04	6.05	3.26	1	0.003732393	TRUE
104	1240.0	Spider	4.22	4.92	4.95	1	0.003733499	TRUE
105	2490.0	Man	3.32	3.95	4.72	1	0.003744393	TRUE
106	2752.0	Alcoholic	4.07	4.3	4.84	1	0.004029394	TRUE
107	3300.0	DisabledChild	2.74	4.55	4.64	1	0.004110516	TRUE
108	7560.0	Freeway	4.47	5.24	4.63	1	0.004241091	TRUE
109	9010.0	BarbedWire	3.94	4.14	4.06	1	0.005028035	TRUE
110	9373.0	Garbage	3.38	5.01	4.86	1	0.005087722	TRUE
111	9190.0	Woman	3.9	3.91	4.89	1	0.005176091	TRUE
112	9469.0	Building	4	4.08	4.92	1	0.005418208	TRUE
113	2750.0	Bum	2.56	4.31	4.48	1	0.005540549	TRUE
114	2682.0	Police	3.69	4.48	4.02	1	0.006096459	TRUE
115	9621.0	Ship	3.22	5.76	3.55	1	0.006211105	TRUE
116	9912.0	Firefighter	3.46	4.68	4.62	1	0.007127932	TRUE
117	6837.0	Police	4.25	4.5	4.43	1	0.008462662	TRUE
118	1390.0	Bees	4.5	5.29	4.75	1	0.008488281	TRUE
119	1274.0	Roaches	3.17	5.39	5.03	1	0.009151602	TRUE
120	6410.0	AimedGun	3.49	5.89	4.29	1	0.009606836	TRUE
121	6370.0	Attack	2.7	6.44	3	1	0.00961449	TRUE
122	9046.0	Family	3.32	4.31	4.61	1	0.009747758	TRUE
123	2525.0	Women	4.06	3.93	5.32	1	0.010018769	TRUE
124	6020.0	ElectricChair	3.41	5.58	4.07	1	0.010508013	TRUE
125	8160.0	RockClimber	5.07	6.97	4.05	1	0.010735899	TRUE
126	9102.0	Heroin	3.34	4.84	4.64	1	0.011061642	TRUE
127	9031.0	Mud	3.01	4.82	4.68	1	0.01139119	TRUE
128	7136.0	CarBoot	3.47	5.01	3.98	1	0.011782882	TRUE
129	6840.0	Police	3.63	5.95	4.72	1	0.011969379	TRUE
130	9402.0	Mob	4.48	5.07	4.85	1	0.012031436	TRUE
131	9480.0	Skull	3.51	5.57	4.56	1	0.012066496	TRUE
132	1111.0	Snakes	3.25	5.2	4.8	1	0.013249888	TRUE
133	2692.0	Bomb	3.36	5.35	3.87	1	0.013331385	TRUE
134	2520.0	ElderlyMan	4.13	4.22	4.44	1	0.013426721	TRUE
135	6836.0	Police	3.46	5.47	4.39	1	0.013632626	TRUE

No	ImageCode	Description	Valence	Arousal	Dominance1	Cluster	Uncertainty	Under75thQuantile
136	9145.0	Cow	3.2	5.05	4.73	1	0.013682157	TRUE
137	9561.0	SickKitty	2.68	4.79	4.53	1	0.015151018	TRUE
138	1220.0	Spider	3.47	5.57	4.54	1	0.015751552	TRUE
139	9160.0	Soldier	3.23	5.87	3.8	1	0.016283477	TRUE
140	6241.0	Gun	3.42	4.54	4.44	1	0.017415509	TRUE
141	6244.0	AimedGun	3.09	5.68	3.43	1	0.018226991	TRUE
142	9622.0	Jet	3.1	6.26	3.66	1	0.018721334	TRUE
143	2400.0	Woman	4.21	4.2	5.02	1	0.020061688	TRUE
144	9186.0	Vultures	3.43	4.88	4.23	1	0.020969417	TRUE
145	2120.0	AngryFace	3.34	5.18	4.52	1	0.022279821	TRUE
146	6210.0	AimedGun	2.95	6.34	3.46	1	0.022906571	TRUE
147	9120.0	OilFires	3.2	5.77	3.79	1	0.024106474	TRUE
148	1230.0	Spider	4.35	4.44	5.09	1	0.025017169	TRUE
149	6561.0	Attack	3.16	4.99	4.58	1	0.02592682	TRUE
150	7497.0	Crowd	5.19	4.97	4.26	1	0.026778764	TRUE
151	7013.0	Lightbulb	4.2	4.11	5.03	1	0.027752751	TRUE
152	2753.0	Alcoholic	3.17	4.29	4.48	1	0.028033484	TRUE
153	1202.0	Spider	3.35	5.94	4.23	1	0.028950371	TRUE
154	9592.0	Injection	3.34	5.23	4.14	1	0.029126308	TRUE
155	2810.0	Boy	4.31	4.47	5.69	1	0.030673603	TRUE
156	6200.0	AimedGun	2.955	6.015	3.42	1	0.030891088	TRUE
157	3181.0	BatteredFem	2.3	5.06	4.31	1	0.031336986	TRUE
158	2276.0	Girl	2.67	4.63	4.4	1	0.033479956	TRUE
159	2280.0	Boy	4.22	3.77	5.7	1	0.035573662	TRUE
160	2271.0	Woman	4.2	3.74	5.66	1	0.036905575	TRUE
161	2700.0	Woman	3.19	4.77	4.44	1	0.037155089	TRUE
162	9941.0	Fire	2.91	5.83	3.28	1	0.037244935	TRUE
163	9411.0	Boy	4.63	5.37	4.91	1	0.039268336	TRUE
164	9584.0	DentalExam	3.34	4.96	3.94	1	0.039703194	TRUE
165	9180.0	Seal	2.99	5.02	4.52	1	0.042593433	TRUE
166	4621.0	Harassment	3.19	4.92	4.37	1	0.050815586	TRUE
167	7135.0	CarDamage	3.17	5.36	3.76	1	0.057252238	TRUE
168	9041.0	ScaredChild	2.98	4.64	4.38	1	0.062953565	TRUE
169	9002.0	Memorial	3.39	4.55	4.03	1	0.063709879	TRUE
170	9571.0	Cat	1.96	5.64	4.17	1	0.064219121	TRUE
171	7290.0	Fish	4.37	3.87	5.85	1	0.068605213	TRUE
172	7632.0	Airplane	5.22	4.78	4.4	1	0.068718937	TRUE
173	7361.0	MeatSlicer	3.1	5.09	4.38	1	0.069362857	TRUE
174	9468.0	Grafitti	4.67	4.68	4.58	1	0.071646974	TRUE
175	1271.0	Roaches	3.19	5.37	4.2	1	0.072581562	TRUE
176	4664.2	Attack	2.79	6.13	3.33	1	0.074179348	TRUE
177	9330.0	Garbage	2.89	4.35	4.33	1	0.078455165	TRUE
178	6220.0	BoysW/Guns	3.1	5.89	3.92	1	0.082156334	TRUE
179	9452.0	Gun	3.19	5.14	4.09	1	0.082671171	TRUE
180	9599.0	Injection	3.16	5.43	4.22	1	0.085310582	TRUE
181	2206.0	Fingerprint	4.06	3.71	4.46	1	0.096223237	TRUE
182	2458.0	CryingBaby	4.69	5.28	5.06	1	0.096418317	TRUE
183	2590.0	ElderlyWoman	3.26	3.93	4.31	1	0.106269288	TRUE
184	9340.0	Garbage	2.41	5.16	4.24	1	0.107810356	FALSE
185	8010.0	Runner	4.38	4.12	5.17	1	0.107843877	FALSE
186	1303.0	Dog	4.68	5.7	4.98	1	0.120602864	FALSE
187	9042.0	StickThruLip	3.15	5.78	4.37	1	0.121022137	FALSE
188	9530.0	Boys	2.93	5.2	4.32	1	0.128496804	FALSE
189	9590.0	Injecting	3.08	5.41	4	1	0.129714804	FALSE
190	1645.0	Wolf	4.99	5.14	4.74	1	0.133315143	FALSE
191	9620.0	Shipwreck	2.7	6.11	3.29	1	0.13506281	FALSE
192	6213.0	Terrorist	2.91	5.86	3.62	1	0.138437588	FALSE
193	9470.0	Ruins	3.05	5.05	4.11	1	0.157849737	FALSE
194	1945.0	Turtle	4.59	4.42	5.57	1	0.159598979	FALSE
195	9007.0	Needles	2.49	5.03	4.18	1	0.161165418	FALSE
196	3180.0	BatteredFem	1.92	5.77	4.05	1	0.162974127	FALSE
197	9415.0	Handicapped	2.82	4.91	4.22	1	0.16888363	FALSE
198	9422.0	Battleship	4.95	5.09	4.89	1	0.187958712	FALSE
199	9417.0	Ticket	3.16	4.83	3.7	1	0.193537143	FALSE
200	2691.0	Riot	3.04	5.85	4.39	1	0.207114967	FALSE
201	9043.0	Teeth	2.52	5.5	4.29	1	0.208284272	FALSE
202	1112.0	Snake	4.71	4.6	5.27	1	0.221319669	FALSE
203	7380.0	RoachOnPizza	2.46	5.88	4.49	1	0.227983706	FALSE
204	6830.0	Guns	2.82	6.21	3.67	1	0.228454579	FALSE
205	2456.0	CryingFamily	2.84	4.55	4.15	1	0.230425585	FALSE
206	3185.0	Stitches	2.81	5.48	4.24	1	0.246199628	FALSE
207	3301.0	InjuredChild	1.8	5.21	3.71	1	0.247698942	FALSE
208	6571.0	CarTheft	2.85	5.59	3.56	1	0.249991211	FALSE
209	2301.0	KidCry	2.78	4.57	4.13	1	0.25147833	FALSE
210	8230.0	Boxer	2.95	5.91	4.05	1	0.25806378	FALSE
211	6834.0	Police	2.91	6.28	3.9	1	0.277087397	FALSE
212	2230.0	SadFace	4.53	4.13	4.8	1	0.281838451	FALSE
213	9430.0	Burial	2.63	5.26	4.14	1	0.284774204	FALSE
214	1820.0	Crocodile	5.35	5.67	4.66	1	0.285596512	FALSE

No	ImageCode	Description	Valence	Arousal	Dominance1	Cluster	Uncertainty	Under75thQuantile
215	9181.0	DeadCows	2.26	5.39	4.04	1	0.290502023	FALSE
216	9909.0	BurningCar	2.78	5.98	3.67	1	0.306723583	FALSE
217	3213.0	Surgery	2.96	6.82	3.92	1	0.31975866	FALSE
218	1908.0	Jellyfish	5.28	4.88	4.75	1	0.331013362	FALSE
219	9280.0	Smoke	2.8	4.26	4.1	1	0.338189577	FALSE
220	9610.0	Accident	2.89	5.23	3.82	1	0.34434321	FALSE
221	9491.0	DeadBody	2.78	5.69	3.64	1	0.347112818	FALSE
222	3160.0	EyeDisease	2.63	5.35	4.08	1	0.358525316	FALSE
223	4233.0	Prostitute	4.56	3.96	5.61	1	0.362733389	FALSE
224	2751.0	DrunkDriving	2.67	5.18	4.01	1	0.391091632	FALSE
225	6540.0	Attack	2.19	6.83	3.02	1	0.399386013	FALSE
226	3019.0	Organs	2.99	6.3	4.25	1	0.416946988	FALSE
227	9321.0	Vomit	2.81	6.24	3.9	1	0.421449916	FALSE
228	4230.0	Prostitute	4.86	4.7	5.38	1	0.424026599	FALSE
229	8121.0	Athlete	4.63	4.14	5.3	1	0.428737309	FALSE
230	9520.0	Kids	2.46	5.41	4.01	1	0.42970857	FALSE
231	2410.0	Boy	4.62	4.13	5	1	0.429775026	FALSE
232	2345.1	BlackEye	2.26	5.5	3.96	1	0.433024973	FALSE
233	9611.0	PlaneCrash	2.71	5.75	3.67	1	0.439357647	FALSE
234	9050.0	PlaneCrash	2.43	6.36	3.27	1	0.443778946	FALSE
235	9905.0	CarAccident	2.55	5.93	3.33	1	0.448759768	FALSE
236	9920.0	CarAccident	2.5	5.76	3.09	1	0.454450219	FALSE
237	2141.0	GrievingFem	2.44	5	3.92	1	0.469382764	FALSE
238	6831.0	Police	2.59	5.55	3.99	1	0.477779854	FALSE
239	9253.0	Mutilation	2	5.53	3.77	1	0.48506186	FALSE
240	9140.0	Cow	2.19	5.38	3.85	1	0.490471023	FALSE
241	9922.0	Fire	2.78	5.21	3.72	1	0.49787033	FALSE
242	4631.0	BikerCouple	5.36	5.19	4.87	1	0.50370197	FALSE
243	7920.0	CarCrash	4.51	3.87	5.2	1	0.511416833	FALSE
244	9401.0	Knives	4.53	3.88	5.29	1	0.517298278	FALSE
1	5301.0	Galaxy	6.54	5.21	4.06	2	0.001073592	TRUE
2	1650.0	Jaguar	6.65	6.23	4.29	2	0.002234877	TRUE
3	2050.0	Baby	8.2	4.57	7.71	2	0.002993239	TRUE
4	1810.0	Hippo	6.52	4.45	4.55	2	0.003845426	TRUE
5	1720.0	Lion	6.79	5.32	4.63	2	0.005560607	TRUE
6	5991.0	Sky	6.55	4.01	4.78	2	0.009872293	TRUE
7	5760.0	Nature	8.05	3.22	7.49	2	0.010131601	TRUE
8	5990.0	Sky	6.54	4.44	4.7	2	0.010612086	TRUE
9	8400.0	Rafters	7.09	6.61	4.63	2	0.013742259	TRUE
10	2260.0	NeutBaby	8.06	4.26	7.47	2	0.014347158	TRUE
11	5870.0	Clouds	6.78	3.1	5.2	2	0.02831353	TRUE
12	8492.0	Rollercoaster	7.21	7.31	4.63	2	0.034153044	TRUE
13	5470.0	Astronaut	7.35	6.02	4.96	2	0.034431135	TRUE
14	2040.0	Baby	8.17	4.64	7.33	2	0.035854044	TRUE
15	5950.0	Lightning	5.99	6.79	3.56	2	0.059391116	TRUE
16	5010.0	Flower	7.14	3	7.4	2	0.090149432	TRUE
17	4700.0	Couple	6.91	4.05	5.35	2	0.126319909	TRUE
18	2360.0	Family	7.7	3.66	6.92	2	0.152611628	TRUE
19	8180.0	CliffDivers	7.12	6.59	4.97	2	0.172624173	TRUE
20	1721.0	Lion	7.3	4.53	5.57	2	0.180106051	TRUE
21	5220.0	Nature	7.01	3.91	5.53	2	0.187465973	TRUE
22	2530.0	Couple	7.8	3.99	5.99	2	0.188125183	TRUE
23	7492.0	Ferry	7.41	4.91	5.55	2	0.191300249	TRUE
24	2540.0	Mother	7.63	3.97	5.96	2	0.204061991	TRUE
25	2510.0	ElderlyWoman	6.91	4	5.46	2	0.207513669	TRUE
26	2501.0	Couple	6.89	3.09	5.63	2	0.208165135	TRUE
27	1610.0	Rabbit	7.755	3.53	6.645	2	0.209582522	TRUE
28	5910.0	Fireworks	7.8	5.59	5.56	2	0.230106902	TRUE
29	5780.0	Nature	7.52	3.75	6.05	2	0.241508238	TRUE
30	5779.0	Courtyard	7.33	3.57	6.96	2	0.26184323	TRUE
31	1750.0	Bunnies	8.28	4.1	6.15	2	0.26563416	TRUE
32	2650.0	Boy	7.27	4.28	5.74	2	0.272959888	TRUE
33	2057.0	Father	7.81	4.54	6.76	2	0.274129359	TRUE
34	4610.0	Romance	7.29	5.1	5.54	2	0.282564743	TRUE
35	2035.0	Kid	7.52	3.69	6.2	2	0.284868598	TRUE
36	5000.0	Flower	7.08	2.67	7.08	2	0.285318025	TRUE
37	1590.0	Horse	7.21	4.77	5.58	2	0.2864767	TRUE
38	5200.0	Flowers	7.36	3.2	6.21	2	0.287443329	TRUE
39	1920.0	Porpoise	7.9	4.27	6.5	2	0.294102492	TRUE
40	2091.0	Girls	7.68	4.51	6.79	2	0.294154118	TRUE
41	1440.0	Seal	8.19	4.61	6.05	2	0.300186824	TRUE
42	1620.0	Antelope	7.37	3.54	6.82	2	0.301887261	TRUE
43	2314.0	Binoculars	7.55	4	6.17	2	0.315425303	TRUE
44	5210.0	Seaside	8.03	4.6	6.19	2	0.327273511	TRUE
45	8090.0	Gymnast	7.02	5.71	5.25	2	0.328009533	TRUE
46	2370.0	ThreeMen	7.14	2.9	6.12	2	0.342472953	TRUE
47	2165.0	Father	7.63	4.55	6.72	2	0.350898665	TRUE
48	2660.0	Baby	7.75	4.44	6.44	2	0.358158048	TRUE
49	2170.0	Mother	7.55	4.08	6.49	2	0.368151907	TRUE

No	ImageCode	Description	Valence	Arousal	Dominance1	Cluster	Uncertainty	Under75thQuantile
50	5829.0	Sunset	7.65	4.68	6.06	2	0.372544397	TRUE
51	2058.0	Baby	7.91	5.09	6.67	2	0.37744521	TRUE
52	4628.0	Wedding	7.23	5.19	5.57	2	0.37892916	TRUE
53	4612.0	Couple	6.82	5.06	5.3	2	0.379152468	TRUE
54	2151.0	Father/Child	7.32	4.37	5.9	2	0.381035061	FALSE
55	2340.0	Family	8.03	4.9	6.18	2	0.383600406	FALSE
56	2550.0	Couple	7.77	4.68	6.22	2	0.384171387	FALSE
57	8370.0	Rafting	7.77	6.73	5.37	2	0.386594005	FALSE
58	1500.0	Dog	7.24	4.12	6.97	2	0.387114352	FALSE
59	5831.0	Seagulls	7.63	4.43	6.46	2	0.397194134	FALSE
60	2311.0	Mother	7.54	4.42	6.16	2	0.399375697	FALSE
61	1600.0	Horse	7.37	4.05	6.75	2	0.401792466	FALSE
62	7570.0	Skyline	6.97	5.54	5.33	2	0.429363638	FALSE
63	2274.0	Kids	7.47	4.22	6.35	2	0.435269515	FALSE
64	2156.0	Family	7.12	4.34	5.82	2	0.466553964	FALSE
65	2341.0	Children	7.38	4.11	6.44	2	0.473667228	FALSE
66	8540.0	Athletes	7.48	5.16	5.88	2	0.475793366	FALSE
67	5202.0	Garden	7.25	3.73	6.31	2	0.480331867	FALSE
68	5480.0	Fireworks	7.53	5.48	5.8	2	0.484569434	FALSE
69	2304.0	Girl	7.22	3.63	6.35	2	0.492653256	FALSE
70	1710.0	Puppies	8.34	5.41	6.55	2	0.493046159	FALSE
71	1811.0	Monkeys	7.62	5.12	6.07	2	0.495840811	FALSE
1	9220.0	Cemetery	2.06	4	3.13	3	0.001106416	TRUE
2	9000.0	Cemetery	2.55	4.06	3.25	3	0.002905535	TRUE
3	9001.0	Cemetery	3.1	3.67	3.47	3	0.010288332	TRUE
4	2205.0	Hospital	1.95	4.53	3.22	3	0.017999846	TRUE
5	3230.0	DyingMan	2.02	5.41	2.93	3	0.05138815	TRUE
6	9421.0	Soldier	2.21	5.04	3.32	3	0.067532115	TRUE
7	3030.0	Mutilation	1.91	6.76	3.69	3	0.071314243	TRUE
8	3130.0	Mutilation	1.58	6.97	3.46	3	0.07249136	TRUE
9	2730.0	NativeBoy	2.45	6.8	3.94	3	0.077951334	TRUE
10	9810.0	KKKrally	2.09	6.62	3.95	3	0.094333693	TRUE
11	3071.0	Mutilation	1.88	6.86	3.28	3	0.10193917	TRUE
12	3120.0	DeadBody	1.56	6.84	3.32	3	0.104260703	TRUE
13	3400.0	SeveredHand	2.35	6.91	3.65	3	0.111773504	TRUE
14	9921.0	Fire	2.04	6.52	3.57	3	0.118612592	TRUE
15	9040.0	StarvingChild	1.67	5.82	3.1	3	0.134295473	TRUE
16	9187.0	InjuredDog	1.81	6.45	3.17	3	0.144685577	TRUE
17	9325.0	Vomit	1.89	6.01	3.22	3	0.145520889	TRUE
18	3140.0	DeadBody	1.83	6.36	3.2	3	0.14598058	TRUE
19	3195.0	Stitches	2.06	6.36	3.55	3	0.147774236	TRUE
20	3100.0	BurnVictim	1.6	6.49	3	3	0.154825601	TRUE
21	2352.2	BloodyKiss	2.09	6.25	3.45	3	0.168181943	TRUE
22	3110.0	BurnVictim	1.79	6.7	3.04	3	0.168913902	TRUE
23	9183.0	HurtDog	1.69	6.58	2.96	3	0.169141937	TRUE
24	3103.0	Injury	2.07	6.06	3.37	3	0.17279913	TRUE
25	9405.0	SlicedHand	1.83	6.08	3.4	3	0.186797657	TRUE
26	6360.0	Attack	2.23	6.33	3.97	3	0.187216009	TRUE
27	3350.0	Infant	1.88	5.72	3.38	3	0.187795359	TRUE
28	6212.0	Soldier	2.19	6.01	3.45	3	0.200877158	TRUE
29	9412.0	DeadMan	1.83	6.72	3	3	0.203116434	TRUE
30	3060.0	Mutilation	1.79	7.12	2.97	3	0.203646087	TRUE
31	9420.0	Soldier	2.31	5.69	3.27	3	0.212646319	TRUE
32	6563.0	Attack	1.77	6.85	2.93	3	0.222131706	TRUE
33	9570.0	Dog	1.68	6.14	3.37	3	0.22377472	TRUE
34	6312.0	Abduction	2.48	6.37	3.83	3	0.228992218	TRUE
35	2053.0	Baby	2.47	5.25	3.56	3	0.229345381	TRUE
36	2900.0	CryingBoy	2.45	5.09	3.64	3	0.233753986	TRUE
37	9911.0	CarAccident	2.3	5.76	3.54	3	0.235740316	TRUE
38	9006.0	HIVTattoo	2.34	5.76	3.33	3	0.246161599	TRUE
39	3150.0	Mutilation	2.26	6.55	3.39	3	0.246818746	TRUE
40	3168.0	Mutilation	1.56	6	3.24	3	0.247610088	TRUE
41	9910.0	CarAccident	2.06	6.2	3.02	3	0.249894348	TRUE
42	9332.0	CryingWoman	2.25	5.34	3.63	3	0.253395258	TRUE
43	2800.0	SadChild	1.78	5.49	3.4	3	0.25592123	TRUE
44	6570.0	Suicide	2.19	6.24	4.03	3	0.260318122	TRUE
45	9414.0	Execution	2.06	6.49	3.11	3	0.261128456	TRUE
46	9560.0	DuckInOil	2.12	5.5	3.62	3	0.273714935	TRUE
47	3550.1	PlaneCrash	2.35	6.29	3.47	3	0.283858117	TRUE
48	3220.0	Hospital	2.49	5.52	3.53	3	0.293524114	TRUE
49	9250.0	WarVictim	2.57	6.6	3.73	3	0.300509285	TRUE
50	3212.0	Surgery	2.79	6.57	4.07	3	0.304435776	TRUE
51	9500.0	Porpoises	2.42	5.82	3.42	3	0.310114665	TRUE
52	9295.0	Garbage	2.39	5.11	3.74	3	0.317974389	TRUE
53	9185.0	DeadDog	1.97	5.65	3.62	3	0.326299083	TRUE
54	6243.0	AimedGun	2.33	5.99	3.23	3	0.331929419	FALSE
55	9400.0	Soldier	2.5	5.99	3.78	3	0.335413046	FALSE
56	9904.0	CarAccident	2.39	6.08	3.4	3	0.338493394	FALSE
57	6838.0	Police	2.45	5.8	3.79	3	0.340064175	FALSE

No	ImageCode	Description	Valence	Arousal	Dominance1	Cluster	Uncertainty	Under75thQuantile
58	9163.0	Soldiers	2.1	6.53	3.04	3	0.363947627	FALSE
59	6560.0	Attack	2.16	6.53	3.11	3	0.3683177	FALSE
60	9184.0	InjuredDog	2.47	5.75	3.86	3	0.378061661	FALSE
61	9432.0	Mastectomy	2.56	4.92	3.83	3	0.387364659	FALSE
62	3550.0	Injury	2.54	5.92	3.64	3	0.389761316	FALSE
63	9322.0	Vomit	2.24	5.73	3.87	3	0.395920654	FALSE
64	9927.0	Flood	2.71	5.29	3.6	3	0.427846739	FALSE
65	6821.0	Gang	2.38	6.29	3.29	3	0.440651181	FALSE
66	9302.0	Toilet	2.32	5.58	3.9	3	0.446040252	FALSE
67	3051.0	Mutilation	2.3	5.62	3.92	3	0.458651054	FALSE
68	6242.0	Gang	2.69	5.43	3.49	3	0.464693208	FALSE
69	6530.0	Attack	2.76	6.18	4.01	3	0.465829671	FALSE
70	9326.0	Vomit	2.21	5.89	4.05	3	0.495433228	FALSE
71	6570.1	Suicide	2.54	6.12	3.46	3	0.499523884	FALSE
1	7010.0	Basket	4.94	1.76	6.7	4	0.008850402	TRUE
2	7004.0	Spoon	5.04	2	6.74	4	0.009448474	TRUE
3	7080.0	Fork	5.27	2.32	7.04	4	0.012002436	TRUE
4	7031.0	Shoes	4.52	2.03	6.14	4	0.013146734	TRUE
5	7217.0	ClothesRack	4.82	2.43	6.25	4	0.015045519	TRUE
6	7175.0	Lamp	4.87	1.72	6.47	4	0.015718893	TRUE
7	7110.0	Hammer	4.55	2.27	6.07	4	0.016139975	TRUE
8	7950.0	Tissue	4.94	2.28	6.3	4	0.017274393	TRUE
9	7090.0	Book	5.19	2.61	6.65	4	0.017283909	TRUE
10	7035.0	Mug	4.98	2.66	6.39	4	0.017831097	TRUE
11	7006.0	Bowl	4.88	2.33	6.18	4	0.019012979	TRUE
12	5510.0	Mushroom	5.15	2.82	6.68	4	0.021727461	TRUE
13	7705.0	Cabinet	4.77	2.65	6.39	4	0.024312502	TRUE
14	7185.0	AbstractArt	4.97	2.64	6.13	4	0.024459873	TRUE
15	7235.0	Chair	4.96	2.83	6.53	4	0.026016116	TRUE
16	5130.0	Rocks	4.45	2.51	5.84	4	0.027960194	TRUE
17	7233.0	Plate	5.09	2.77	6.23	4	0.028102958	TRUE
18	7491.0	Building	4.82	2.39	5.93	4	0.029920612	TRUE
19	7000.0	RollingPin	5	2.42	6.14	4	0.030622358	TRUE
20	7025.0	Stool	4.63	2.71	6.1	4	0.030888533	TRUE
21	2190.0	Man	4.83	2.41	5.92	4	0.031206898	TRUE
22	2411.0	Girl	5.07	2.86	6.15	4	0.031324572	TRUE
23	2890.0	Twins	4.95	2.95	5.99	4	0.032420648	TRUE
24	7060.0	TrashCan	4.43	2.55	5.85	4	0.032561141	TRUE
25	7012.0	Rubberbands	4.98	3	6.06	4	0.032703311	TRUE
26	7009.0	Mug	4.93	3.01	6.33	4	0.034453437	TRUE
27	2570.0	Man	4.78	2.76	5.72	4	0.034808634	TRUE
28	2440.0	NeutGirl	4.49	2.63	5.97	4	0.036238652	TRUE
29	7003.0	Disk	5	3.07	6.02	4	0.037140054	TRUE
30	7020.0	Fan	4.97	2.17	6.16	4	0.03730284	TRUE
31	7030.0	Iron	4.69	2.99	5.73	4	0.040477181	TRUE
32	7050.0	HairDryer	4.93	2.75	5.82	4	0.042273915	TRUE
33	7002.0	Towel	4.97	3.16	6.25	4	0.042539004	TRUE
34	5534.0	Mushrooms	4.84	3.14	5.83	4	0.043319719	TRUE
35	5520.0	Mushroom	5.33	2.95	6.43	4	0.045213537	TRUE
36	5740.0	Plant	5.21	2.59	6.27	4	0.045390845	TRUE
37	7034.0	Hammer	4.95	3.06	6.48	4	0.045655566	TRUE
38	2381.0	Girl	5.25	3.04	6.28	4	0.046165738	TRUE
39	7032.0	Shoes	4.82	3.18	5.9	4	0.047751403	TRUE
40	7014.0	Scissors	5.15	3.25	6.21	4	0.048912648	TRUE
41	7150.0	Umbrella	4.72	2.61	5.55	4	0.049213587	TRUE
42	7040.0	DustPan	4.69	2.69	5.46	4	0.05248799	TRUE
43	7160.0	Fabric	5.02	3.07	5.8	4	0.052682763	TRUE
44	2002.0	Man	4.95	3.35	5.89	4	0.057468026	TRUE
45	5530.0	Mushroom	5.38	2.87	6.42	4	0.057758337	TRUE
46	7255.0	Cracker	5.07	3.36	5.92	4	0.060585468	TRUE
47	9700.0	Trash	4.77	3.21	5.47	4	0.061725983	TRUE
48	2880.0	Shadow	5.18	2.96	6.01	4	0.062044169	TRUE
49	7019.0	Tools	5.2	3.36	6.19	4	0.062858766	TRUE
50	2200.0	NeutFace	4.79	3.18	5.44	4	0.063036654	TRUE
51	7187.0	AbstractArt	5.07	2.3	6.1	4	0.064265332	TRUE
52	5500.0	Mushroom	5.42	3	6.45	4	0.064943909	TRUE
53	7170.0	LightBulb	5.14	3.21	5.89	4	0.065886379	TRUE
54	7045.0	Zipper	4.97	3.32	6.28	4	0.06669553	TRUE
55	2214.0	NeutMan	5.01	3.46	5.98	4	0.069724741	TRUE
56	7017.0	Video	5.18	3.12	5.93	4	0.070596476	TRUE
57	2870.0	Teenager	5.31	3.01	6.17	4	0.074573231	TRUE
58	2516.0	ElderlyWoman	4.9	3.5	5.54	4	0.083554984	TRUE
59	7016.0	Razor	4.76	3.4	5.74	4	0.084649188	TRUE
60	5533.0	Mushrooms	5.31	3.12	6.09	4	0.08756537	TRUE
61	7207.0	Beads	5.15	3.57	6	4	0.088712254	TRUE
62	7180.0	NeonBuilding	4.73	3.43	5.61	4	0.092602096	TRUE
63	2514.0	Woman	5.19	3.5	5.85	4	0.094065433	TRUE
64	9070.0	Boy	5.01	3.63	5.67	4	0.095819366	TRUE
65	2383.0	Secretary	4.72	3.41	5.75	4	0.100191421	TRUE

No	ImageCode	Description	Valence	Arousal	Dominance1	Cluster	Uncertainty	Under75thQuantile
66	7062.0	Sewing	5.27	3.4	5.91	4	0.106009206	TRUE
67	2850.0	Tourist	5.22	3	5.87	4	0.107415035	TRUE
68	6150.0	Outlet	5.08	3.22	5.54	4	0.109172681	TRUE
69	2385.0	Girl	5.2	3.64	5.86	4	0.111142448	TRUE
70	2480.0	ElderlyMan	4.77	2.66	5.33	4	0.112492643	TRUE
71	5531.0	Mushroom	5.15	3.69	6	4	0.112928617	TRUE
72	2377.0	Reading	5.19	3.5	5.68	4	0.117168969	TRUE
73	7100.0	FireHydrant	5.24	2.89	5.92	4	0.117868204	TRUE
74	2495.0	Man	5.22	3.19	5.77	4	0.118781053	TRUE
75	7184.0	AbstractArt	4.84	3.66	5.46	4	0.120432578	TRUE
76	7700.0	Office	4.25	2.95	5.13	4	0.120958627	TRUE
77	2215.0	NeutMan	4.63	3.38	5.68	4	0.122281002	TRUE
78	2221.0	Judge	4.39	3.07	4.97	4	0.122352188	TRUE
79	7287.0	Tomato	4.77	3.57	5.68	4	0.122983467	TRUE
80	2210.0	NeutFace	4.54	3.32	5.13	4	0.127835768	TRUE
81	2484.0	Amerindian	5	3.75	5.45	4	0.128073173	TRUE
82	7130.0	Truck	4.77	3.35	5.08	4	0.130668702	TRUE
83	2026.0	Woman	4.82	3.4	5.09	4	0.140113604	TRUE
84	5532.0	Mushrooms	5.19	3.79	6.01	4	0.144400649	TRUE
85	9360.0	EmptyPool	4.03	2.63	5.34	4	0.147352538	TRUE
86	2830.0	Woman	4.73	3.64	5.33	4	0.148483434	TRUE
87	2749.0	Smoking	5.04	3.76	5.35	4	0.148499391	TRUE
88	5120.0	PineNeedles	4.39	3.07	5.69	4	0.155067623	TRUE
89	7026.0	PicnicTable	5.38	2.63	6.19	4	0.165460169	TRUE
90	9260.0	Hands	4.63	3.45	4.98	4	0.171522964	TRUE
91	7001.0	Buttons	5.32	3.2	5.82	4	0.172941304	TRUE
92	7205.0	Scarves	5.56	2.93	6.39	4	0.173528799	TRUE
93	2840.0	Chess	4.91	2.43	5.56	4	0.176848159	TRUE
94	8312.0	Golf	5.37	3.32	5.88	4	0.177717624	TRUE
95	2702.0	BingeEating	5.21	3.92	5.7	4	0.17817781	TRUE
96	5731.0	Flowers	5.39	2.74	6.13	4	0.181986523	TRUE
97	7710.0	Bed	5.42	3.44	5.96	4	0.186108945	TRUE
98	7224.0	FileCabinets	4.45	2.81	6.26	4	0.186127072	TRUE
99	2308.0	GirlMakeup	5.22	3.82	5.5	4	0.186577362	TRUE
100	1616.0	Bird	5.21	3.95	5.67	4	0.187330342	TRUE
101	2390.0	Couple	5.4	3.57	5.89	4	0.194116455	TRUE
102	7061.0	Puzzle	5.4	3.66	5.92	4	0.194325563	TRUE
103	6570.2	BlowDry	4.86	3.85	5.66	4	0.194335934	TRUE
104	4571.0	AttractiveMan	5.49	3.54	6.13	4	0.19929769	TRUE
105	7182.0	Checkerboard	5.16	4.02	5.51	4	0.201770807	TRUE
106	2101.0	Man	4.49	3.46	5.25	4	0.202587471	TRUE
107	7512.0	Chess	5.38	3.72	5.84	4	0.204629065	TRUE
108	2720.0	Urinating	5.43	3.43	5.92	4	0.211121735	TRUE
109	2279.0	Braces	4.71	3.74	5.55	4	0.219889136	TRUE
110	2487.0	Musician	5.2	4.05	5.81	4	0.232683627	TRUE
111	5040.0	Venusflytrap	5.39	3.75	5.77	4	0.232862319	TRUE
112	7354.0	Garlic	5.51	3.73	6.27	4	0.234245162	TRUE
113	7590.0	Traffic	4.75	3.8	5.05	4	0.241998055	TRUE
114	4000.0	Artist	4.82	3.97	5.35	4	0.250379826	TRUE
115	7081.0	Luggage	5.36	3.96	5.76	4	0.253519409	FALSE
116	5390.0	Boat	5.59	2.88	6.33	4	0.268367083	FALSE
117	2211.0	Man	5.19	4.05	5.25	4	0.269105274	FALSE
118	7237.0	AbstractArt	5.43	3.88	5.84	4	0.27475933	FALSE
119	7513.0	Crochet	5.45	3.47	5.82	4	0.280392388	FALSE
120	7830.0	Agate	5.26	4.08	5.36	4	0.281053682	FALSE
121	7283.0	Fruit	5.5	3.81	6.05	4	0.281953074	FALSE
122	7365.0	Meat	5.2	4.13	5.83	4	0.283844258	FALSE
123	7018.0	Screw	4.81	3.91	5.71	4	0.287849053	FALSE
124	7234.0	IroningBoard	4.23	2.96	5.73	4	0.307058152	FALSE
125	2107.0	Male	5.53	3.72	5.92	4	0.333735279	FALSE
126	7550.0	Office	5.27	3.95	5.22	4	0.335371559	FALSE
127	7186.0	AbstractArt	4.63	3.6	5.88	4	0.341959995	FALSE
128	9210.0	Rain	4.53	3.08	4.55	4	0.343202083	FALSE
129	1122.0	Lizard	5.15	4.32	5.55	4	0.345409485	FALSE
130	7300.0	Peanuts	5.64	3.25	6.2	4	0.366099577	FALSE
131	7595.0	Traffic	4.55	3.77	5.28	4	0.372101993	FALSE
132	2690.0	Terrorist	4.78	4.02	4.91	4	0.394452208	FALSE
133	1350.0	Pig	5.25	4.37	5.6	4	0.397807426	FALSE
134	2372.0	Woman	5.48	4.09	5.72	4	0.407018604	FALSE
135	4274.0	AttractiveFem	5.42	4.18	5.47	4	0.414202374	FALSE
136	5395.0	Boat	5.34	4.23	5.23	4	0.41780329	FALSE
137	7096.0	Car	5.54	3.98	5.81	4	0.429451522	FALSE
138	7285.0	Tomatoes	5.67	3.83	6.29	4	0.430578095	FALSE
139	5535.0	Stilllife	4.81	4.11	5.61	4	0.431184735	FALSE
140	2032.0	Makeup	5.58	4	6.14	4	0.433991823	FALSE
141	7011.0	GasCan	4.52	3.81	4.99	4	0.434475937	FALSE
142	7033.0	Train	5.4	3.99	5.32	4	0.436027373	FALSE
143	7820.0	Agate	5.39	4.21	5.3	4	0.439506463	FALSE
144	2575.0	Propeller	5.46	4.16	6.11	4	0.45369198	FALSE

No	ImageCode	Description	Valence	Arousal	Dominance1	Cluster	Uncertainty	Under75thQuantile
145	4573.0	AttractiveMale	5.49	3.96	5.54	4	0.456100952	FALSE
146	7211.0	Clock	4.81	4.2	4.99	4	0.458154929	FALSE
147	2309.0	GirlCow	4.89	4.33	5.39	4	0.467166493	FALSE
148	4605.0	Couple	5.59	3.84	5.83	4	0.474813621	FALSE
149	2485.0	Man	5.69	3.74	6.12	4	0.49230731	FALSE
150	7021.0	Whistle	5.21	4.17	6.22	4	0.50896079	FALSE
151	4004.0	EroticFemale	5.14	4.44	5.75	4	0.544323125	FALSE
152	2122.0	TongueOut	5.15	4.59	5.49	4	0.553508061	FALSE
1	4232.0	EroticFemale	5.95	6.28	5.69	5	0.00011441	TRUE
2	7240.0	Gym	6.02	5.51	6.37	5	0.000248908	TRUE
3	4300.0	EroticFemale	5.7	5.99	5.55	5	0.000606442	TRUE
4	4672.0	EroticCouple	6	6.29	5.38	5	0.000767947	TRUE
5	4693.0	EroticCouple	6.16	6.57	5.46	5	0.000938276	TRUE
6	4651.0	EroticCouple	6.32	6.34	5.8	5	0.001126242	TRUE
7	4683.0	EroticCouple	6.17	6.62	5.39	5	0.001337013	TRUE
8	4669.0	EroticCouple	5.97	6.11	5.34	5	0.001367096	TRUE
9	4692.0	EroticCouple	5.87	6.39	5.15	5	0.001390269	TRUE
10	8117.0	Hockey	6.02	5.3	6.07	5	0.001851148	TRUE
11	4800.0	EroticCouple	6.44	7.07	5.51	5	0.002318753	TRUE
12	4666.0	EroticCouple	6.24	6.1	5.55	5	0.002512845	TRUE
13	4310.0	EroticFemale	6.04	5.42	5.77	5	0.002519993	TRUE
14	4505.0	AttractiveMale	6.21	5.52	5.76	5	0.002530742	TRUE
15	4664.0	EroticCouple	6.61	6.72	5.96	5	0.002859151	TRUE
16	4698.0	EroticCouple	6.5	6.72	5.7	5	0.00293259	TRUE
17	9156.0	Plane	6.43	5.79	6.04	5	0.002968012	TRUE
18	7279.0	Alcohol	6.22	5.19	5.97	5	0.003002176	TRUE
19	4604.0	EroticCouple	5.98	6.09	5.21	5	0.0034662	TRUE
20	5622.0	Shark	6.33	5.34	5.94	5	0.003817661	TRUE
21	4535.0	Weightlifter	6.27	4.95	6.2	5	0.003835107	TRUE
22	4668.0	EroticCouple	6.67	7.13	5.73	5	0.003868514	TRUE
23	4085.0	EroticFemale	5.71	5.77	5.43	5	0.004105536	TRUE
24	7289.0	Food	6.32	5.14	6.02	5	0.004135842	TRUE
25	4533.0	AttractiveMan	6.22	5.01	5.91	5	0.005032595	TRUE
26	5628.0	Mountains	6.51	5.46	6.28	5	0.005597103	TRUE
27	4658.0	EroticCouple	6.62	6.47	5.86	5	0.005624331	TRUE
28	4697.0	EroticCouple	6.22	6.62	5.21	5	0.005655443	TRUE
29	4007.0	AttractiveFem	6.26	5.63	5.57	5	0.005895486	TRUE
30	4653.0	EroticCouple	6.56	5.83	6.07	5	0.006142509	TRUE
31	4008.0	EroticFemale	5.91	5.66	5.34	5	0.00654752	TRUE
32	4656.0	EroticCouple	6.73	6.41	6.1	5	0.007238519	TRUE
33	8220.0	Runners	6.5	5.19	6.28	5	0.007330674	TRUE
34	7477.0	Sushi	6.12	4.82	6.25	5	0.007471746	TRUE
35	4652.0	EroticCouple	6.79	6.62	6.1	5	0.007501411	TRUE
36	4611.0	EroticCouple	6.62	6.04	5.99	5	0.007727901	TRUE
37	8060.0	Boxer	5.36	5.31	5.92	5	0.007753321	TRUE
38	7440.0	Cookout	6.32	4.7	5.98	5	0.008659384	TRUE
39	8118.0	Rugby	6.14	4.9	5.77	5	0.008841083	TRUE
40	8250.0	Motorcyclist	6.19	5.04	5.63	5	0.009486356	TRUE
41	7461.0	FrenchFries	5.8	5.2	5.88	5	0.009730296	TRUE
42	8280.0	Diver	6.38	5.05	5.85	5	0.010080757	TRUE
43	7281.0	Food	6.4	4.41	6.46	5	0.010611379	TRUE
44	8260.0	Motorcyclist	6.18	5.85	5.29	5	0.010741639	TRUE
45	7650.0	City	6.62	6.15	5.79	5	0.011135755	TRUE
46	5626.0	HangGlider	6.71	6.1	6	5	0.011725116	TRUE
47	7475.0	Shrimp	6.33	4.17	6.34	5	0.012139363	TRUE
48	7402.0	Pastry	5.98	5.05	5.75	5	0.012536915	TRUE
49	4325.0	AttractiveFem	5.96	5.18	5.58	5	0.012664653	TRUE
50	7450.0	Cheeseburger	6.4	5.05	5.81	5	0.012921381	TRUE
51	7505.0	Cards	6.1	4.72	5.95	5	0.013146916	TRUE
52	4606.0	Romance	6.55	5.11	6.09	5	0.014097422	TRUE
53	7286.0	Pancakes	6.36	4.44	5.97	5	0.014189444	TRUE
54	2374.0	Woman	6.29	3.86	6.21	5	0.014288896	TRUE
55	6250.2	IceCream	6.32	5.13	5.63	5	0.014415188	TRUE
56	4687.0	EroticCouple	6.87	6.51	6.04	5	0.014462735	TRUE
57	8065.0	Kickboxing	5.25	5.71	5.52	5	0.014729955	TRUE
58	4090.0	Bikini	6.17	5.39	5.37	5	0.015135225	TRUE
59	2056.0	Diving	6.34	4.63	5.83	5	0.015158895	TRUE
60	8161.0	HangGlider	6.71	6.09	5.89	5	0.015261084	TRUE
61	8032.0	IceSkater	6.38	4.19	6.1	5	0.01552864	TRUE
62	2389.0	Teens	6.61	5.63	5.9	5	0.015886995	TRUE
63	4617.0	EroticFemale	6.6	5.19	6.13	5	0.016044352	TRUE
64	1942.0	Turtles	6.26	4.01	5.95	5	0.016046232	TRUE
65	4532.0	AttractiveMan	6.4	4.15	6.16	5	0.016073049	TRUE
66	2010.0	Adult	6.25	3.32	6.24	5	0.016396764	TRUE
67	4810.0	EroticCouple	6.56	6.66	5.41	5	0.016524836	TRUE
68	8041.0	Diver	6.65	5.49	6.05	5	0.017001827	TRUE
69	1850.0	Camels	6.15	4.06	5.94	5	0.018176883	TRUE
70	8158.0	Hiker	6.53	6.49	5.41	5	0.018324315	TRUE
71	8050.0	Rower	6.24	4.31	6.67	5	0.018693996	TRUE

No	ImageCode	Description	Valence	Arousal	Dominance1	Cluster	Uncertainty	Under75thQuantile
72	2620.0	Woman	5.93	2.72	6.11	5	0.019242574	TRUE
73	2270.0	NeutChild	6.28	3.15	6.49	5	0.019291261	TRUE
74	2392.0	ManW/Fish	6.15	3.85	6.03	5	0.019329421	TRUE
75	4659.0	EroticCouple	6.87	6.93	5.67	5	0.019351295	TRUE
76	4690.0	EroticCouple	6.83	6.06	6.12	5	0.019380823	TRUE
77	5665.0	Building	6.15	4.02	5.78	5	0.019989924	TRUE
78	4100.0	MaleDancers	6.11	4.39	5.93	5	0.019996415	TRUE
79	7600.0	Dragon	5.9	5.5	5.22	5	0.020249502	TRUE
80	7451.0	Hamburger	6.68	5.84	5.85	5	0.020430591	TRUE
81	8031.0	Skier	6.76	5.58	6.36	5	0.02045284	TRUE
82	7481.0	Food	6.53	4.92	5.97	5	0.020604963	TRUE
83	2217.0	Class	6.24	4.08	5.8	5	0.020636022	TRUE
84	5726.0	Grain	6.23	2.84	6.15	5	0.02076035	TRUE
85	7507.0	Painting	6.25	3.54	5.98	5	0.020777275	TRUE
86	2342.0	Children	6.2	4.06	5.76	5	0.021204407	TRUE
87	7515.0	Crowd	6.19	5.48	5.3	5	0.021269112	TRUE
88	1560.0	Hawk	5.97	5.51	5.18	5	0.021499495	TRUE
89	2616.0	Dancer	5.97	4.96	5.58	5	0.021574302	TRUE
90	1595.0	Pony	6.22	4.79	5.54	5	0.021606535	TRUE
91	2092.0	Clowns	6.28	4.32	5.75	5	0.021912465	TRUE
92	4681.0	EroticCouple	6.69	6.68	5.5	5	0.02208943	TRUE
93	7238.0	AbstractArt	6.43	4.17	6.06	5	0.022188917	TRUE
94	8116.0	Football	6.82	5.97	6.06	5	0.022574324	TRUE
95	7352.0	Pizza	6.2	4.58	5.58	5	0.022626187	TRUE
96	7320.0	Desserts	6.54	4.44	6.22	5	0.023204661	TRUE
97	2630.0	Male	6.35	3.92	5.97	5	0.023646765	TRUE
98	4609.0	Couple	6.71	5.54	6	5	0.024643895	TRUE
99	2034.0	Cheerleaders	5.9	4.93	5.79	5	0.025773689	TRUE
100	7284.0	Fruit	6.21	4.06	6.45	5	0.026783596	TRUE
101	4670.0	EroticCouple	6.99	6.74	5.85	5	0.026845225	TRUE
102	2560.0	Picnic	6.34	3.49	6.06	5	0.026981827	TRUE
103	2019.0	AttractiveFem	6.07	4.31	5.89	5	0.02704366	TRUE
104	5020.0	Flower	6.32	2.63	6.67	5	0.027378027	TRUE
105	4071.0	AttractiveFem	5.97	5.14	5.31	5	0.027981253	TRUE
106	5410.0	Violinist	6.11	3.29	6.28	5	0.028206745	TRUE
107	8033.0	IceSkater	6.66	5.01	6.12	5	0.028352213	TRUE
108	8001.0	Basketball	6.5	5.6	5.57	5	0.02847051	TRUE
109	1670.0	Cow	6.315	3.19	6.08	5	0.029075506	TRUE
110	7195.0	Teeth	6.02	4.54	5.78	5	0.029299344	TRUE
111	2302.0	ChildCamera	6.43	3.64	6.18	5	0.029569136	TRUE
112	2511.0	Woman	6.21	3.41	5.81	5	0.031560093	TRUE
113	4608.0	EroticCouple	7.07	6.47	6.25	5	0.031696451	TRUE
114	4607.0	EroticCouple	7.03	6.34	6.14	5	0.033838638	TRUE
115	1450.0	Gannet	6.37	2.83	6.75	5	0.033973526	TRUE
116	4001.0	EroticFemale	5.24	5.24	5.74	5	0.03416971	TRUE
117	4320.0	EroticFemale	6.01	5.11	5.24	5	0.036549621	TRUE
118	4619.0	Romance	6.46	5.09	5.62	5	0.036965313	TRUE
119	8034.0	Skier	7.06	6.3	6.26	5	0.037423973	TRUE
120	7496.0	Street	5.92	4.84	5.55	5	0.037700487	TRUE
121	1602.0	Butterfly	6.5	3.43	6.41	5	0.039438205	TRUE
122	8206.0	Surfers	6.43	6.41	5.19	5	0.039620543	TRUE
123	5875.0	Bicyclist	6.03	3.29	6.19	5	0.041085204	TRUE
124	2320.0	Girl	6.17	2.9	6.66	5	0.041526433	TRUE
125	2240.0	NeutChild	6.53	3.75	6.23	5	0.042741454	TRUE
126	2000.0	Adult	6.51	3.32	6.65	5	0.045293529	TRUE
127	8021.0	Skier	6.79	5.67	5.85	5	0.045813084	TRUE
128	7165.0	BathRoom	6.09	3.5	6.3	5	0.047739169	TRUE
129	7282.0	Cake	6.72	4.77	6.16	5	0.047843447	TRUE
130	4601.0	Romance	6.82	5.08	6.37	5	0.047930475	TRUE
131	1121.0	Lizard	5.79	4.83	5.89	5	0.050558482	TRUE
132	5849.0	Flowers	6.65	4.89	5.9	5	0.050825227	TRUE
133	7489.0	Ferrry	6.54	4.49	5.85	5	0.051885306	TRUE
134	5250.0	Nature	6.08	3.64	5.5	5	0.052670252	TRUE
135	4575.0	AttractiveMale	6.49	4.82	5.66	5	0.053121076	TRUE
136	8320.0	CarRacer	6.24	4.27	5.51	5	0.053759506	TRUE
137	5593.0	Sky	6.47	3.98	5.89	5	0.054545321	TRUE
138	2500.0	Man	6.16	3.61	5.57	5	0.054852949	TRUE
139	7490.0	Window	5.52	2.42	5.81	5	0.056529746	TRUE
140	4530.0	EroticMale	6.19	5.31	5.19	5	0.056803793	TRUE
141	2580.0	Chess	5.71	2.79	5.88	5	0.05851313	TRUE
142	4470.0	EroticMale	5.87	4.81	5.45	5	0.0586365	TRUE
143	1659.0	Gorilla	6.57	4.89	5.71	5	0.062267503	TRUE
144	2250.0	NeutBaby	6.64	4.19	6.85	5	0.063095157	TRUE
145	5215.0	Harbor	6.83	5.4	5.92	5	0.064725614	TRUE
146	8340.0	Plane	6.85	5.8	5.77	5	0.065365713	TRUE
147	7499.0	Concert	6.47	5.58	5.37	5	0.066546393	TRUE
148	2155.0	Pregnant	6.78	5.43	5.81	5	0.066602464	TRUE
149	4536.0	AttractiveMan	6.01	3.95	6.09	5	0.067802511	TRUE
150	7900.0	Violin	6.5	2.6	6.48	5	0.068222637	TRUE

No	ImageCode	Description	Valence	Arousal	Dominance1	Cluster	Uncertainty	Under75thQuantile
151	5629.0	Hiker	7.03	6.55	5.68	5	0.068235324	TRUE
152	2791.0	Balloons	6.64	3.83	6.25	5	0.069673019	TRUE
153	8311.0	Golfer	5.88	3.57	5.8	5	0.069812955	TRUE
154	7501.0	City	6.85	5.63	5.82	5	0.069966635	TRUE
155	5994.0	Skyline	6.8	4.61	6.56	5	0.070507069	TRUE
156	2359.0	Mother/Child	5.87	3.94	5.49	5	0.072740061	TRUE
157	2303.0	Children	6.83	5.53	5.81	5	0.074469882	TRUE
158	7095.0	Headlight	5.99	4.21	6.17	5	0.075670988	TRUE
159	4490.0	EroticMale	6.27	6.06	5.03	5	0.075821408	TRUE
160	2344.0	Children	6.72	4.71	5.96	5	0.076060907	TRUE
161	8465.0	Runner	5.96	3.93	5.97	5	0.078279612	TRUE
162	7509.0	Paintbrush	6.03	3.43	6.31	5	0.080428192	TRUE
163	4520.0	EroticMale	6.6	5.14	5.605	5	0.081378147	TRUE
164	4130.0	EroticFemales	5.36	5.15	5.57	5	0.083422292	TRUE
165	4003.0	EroticFemale	5.48	5.09	5.59	5	0.083686431	TRUE
166	4689.0	EroticCouple	6.9	6.21	5.6	5	0.084553394	TRUE
167	8208.0	Surfer	6.79	5.17	5.85	5	0.084978014	TRUE
168	2352.0	Kiss	6.94	4.99	6.32	5	0.092135297	TRUE
169	7140.0	Bus	5.5	2.92	5.45	5	0.095161288	TRUE
170	2060.0	Baby	6.49	3.8	5.81	5	0.096111998	TRUE
171	5750.0	Nature	6.6	3.14	6.82	5	0.097071569	TRUE
172	4599.0	Romance	7.12	5.69	6.49	5	0.097442492	TRUE
173	2515.0	Harvest	6.09	3.8	6.52	5	0.098264703	TRUE
174	8600.0	Mascot	6.38	4.26	5.54	5	0.098826427	TRUE
175	8500.0	Gold	6.96	5.6	5.87	5	0.101888625	TRUE
176	8130.0	PoleVaulter	6.58	5.49	5.43	5	0.102194035	TRUE
177	4613.0	Condom	5.34	4.66	6.16	5	0.102542551	TRUE
178	8163.0	Parachute	7.14	6.53	5.69	5	0.102649616	TRUE
179	8162.0	HotAirBalloon	6.97	4.98	6.37	5	0.103616079	TRUE
180	4680.0	EroticCouple	7.25	6.02	6.27	5	0.103860849	TRUE
181	7660.0	Crowd	6.61	5.59	5.42	5	0.109334232	TRUE
182	7530.0	House	6.71	4	6.09	5	0.11000901	TRUE
183	5030.0	Flower	6.51	2.74	7.03	5	0.111345485	TRUE
184	4660.0	EroticCouple	7.4	6.58	5.96	5	0.113074393	TRUE
185	2900.2	SmilingGirl	6.62	4.52	5.73	5	0.115939532	TRUE
186	1605.0	Butterfly	6.59	3.43	6.02	5	0.119448218	TRUE
187	7460.0	FrenchFries	6.81	5.12	5.78	5	0.121456659	TRUE
188	5455.0	Cockpit	5.79	4.56	5.63	5	0.121986869	TRUE
189	5623.0	Windsurfers	7.19	5.67	6.45	5	0.123968582	TRUE
190	5900.0	Desert	5.93	4.38	5.16	5	0.12502639	TRUE
191	1910.0	Grouper	6.71	3.29	6.44	5	0.126606747	TRUE
192	2655.0	Child	6.88	4.57	6.14	5	0.126794725	TRUE
193	8531.0	SportCar	7.03	5.41	6.77	5	0.128736401	TRUE
194	1722.0	Jaguars	7.04	5.22	6.12	5	0.130093373	TRUE
195	5621.0	SkyDivers	7.57	6.99	5.81	5	0.132288684	TRUE
196	4597.0	Romance	6.95	5.91	5.64	5	0.133748326	TRUE
197	7390.0	IceCream	6.84	4.56	6.02	5	0.133817811	TRUE
198	1900.0	Fish	6.65	3.46	6.07	5	0.137838517	TRUE
199	2521.0	ManW/Dog	5.78	4.1	5.43	5	0.138534994	TRUE
200	2300.0	AttractiveFem	7.04	5.55	5.89	5	0.139441892	TRUE
201	7400.0	Candy	7	5.06	6.07	5	0.141429575	TRUE
202	7220.0	Pastry	6.91	5.3	5.8	5	0.141880255	TRUE
203	7495.0	Store	5.9	3.82	6.04	5	0.144229736	TRUE
204	2600.0	Beer	5.84	4.16	5.84	5	0.14895508	TRUE
205	2488.0	Musician	5.73	3.91	5.4	5	0.159396149	TRUE
206	4150.0	AttractiveFem	6.53	4.86	5.45	5	0.161398121	TRUE
207	8200.0	WaterSkier	7.54	6.35	6.17	5	0.161525259	TRUE
208	2018.0	VeiledWoman	5.56	4.92	5.5	5	0.164141888	TRUE
209	4600.0	Romance	6.41	4.83	5.33	5	0.166011591	TRUE
210	6910.0	Bomber	5.31	5.62	5.1	5	0.167935691	TRUE
211	7503.0	CardDealer	5.77	4.21	5.59	5	0.17447622	TRUE
212	4640.0	Romance	7.18	5.52	6.03	5	0.174692012	TRUE
213	4750.0	NudeFemale	5.57	4.9	5.48	5	0.17539004	TRUE
214	7620.0	Jet	5.78	4.92	5.07	5	0.175616764	TRUE
215	7351.0	Pizza	5.82	4.25	6	5	0.177342597	TRUE
216	1660.0	Gorilla	6.49	4.57	5.46	5	0.177556158	TRUE
217	1601.0	Giraffes	6.86	3.92	6.24	5	0.177638915	TRUE
218	5199.0	Garden	6.93	4.7	5.99	5	0.177756864	TRUE
219	1640.0	Coyote	6.215	5.155	5.065	5	0.183947788	TRUE
220	8232.0	Boxer	5.07	5.1	5.57	5	0.186042591	TRUE
221	4460.0	EroticMale	5.6	4.94	5.34	5	0.188149599	TRUE
222	8120.0	Athlete	7.09	4.85	6.23	5	0.192963495	TRUE
223	2208.0	Bride	7.35	5.68	6.21	5	0.197549219	TRUE
224	4603.0	Romance	7.1	4.89	6.2	5	0.197634994	TRUE
225	4641.0	Romance	7.2	5.43	6.01	5	0.208269482	TRUE
226	5800.0	Leaves	6.36	2.51	5.72	5	0.209678591	TRUE
227	8040.0	Diver	6.64	5.61	5.31	5	0.209956445	TRUE
228	8080.0	Sailing	7.73	6.65	5.91	5	0.210452165	TRUE
229	2384.0	Fisherman	5.92	3.41	6.32	5	0.211502064	TRUE

No	ImageCode	Description	Valence	Arousal	Dominance1	Cluster	Uncertainty	Under75thQuantile
230	7350.0	Pizza	7.1	4.97	6.72	5	0.21295369	TRUE
231	4005.0	EroticFemale	5.43	5.02	5.39	5	0.216123852	TRUE
232	2036.0	Woman	5.8	3.24	6.1	5	0.217271808	TRUE
233	7410.0	Candy	6.91	4.55	5.92	5	0.220013702	TRUE
234	1510.0	Dog	7.01	4.28	6.29	5	0.222156586	FALSE
235	5720.0	Farmland	6.31	2.79	5.58	5	0.234598143	FALSE
236	4275.0	AttractiveFem	5.7	4.41	5.67	5	0.23476	FALSE
237	8185.0	Skydivers	7.57	7.27	5.47	5	0.235311502	FALSE
238	7545.0	Ocean	6.84	3.28	6.75	5	0.241004455	FALSE
239	4650.0	EroticCouple	6.96	5.67	5.56	5	0.242504951	FALSE
240	7230.0	Turkey	7.38	5.52	6.21	5	0.243331293	FALSE
241	2518.0	Quilting	5.67	3.31	5.8	5	0.245582078	FALSE
242	2075.0	Baby	7.32	5.27	6.42	5	0.249237549	FALSE
243	8502.0	Money	7.51	5.78	6.4	5	0.250673773	FALSE
244	1603.0	Butterfly	6.9	3.37	6.57	5	0.255531863	FALSE
245	8501.0	Money	7.91	6.44	6.05	5	0.258318233	FALSE
246	1340.0	Women	7.13	4.75	6.13	5	0.260192732	FALSE
247	4500.0	AttractiveMan	5.7	3.68	5.72	5	0.266107801	FALSE
248	8470.0	Gymnast	7.74	6.14	6.17	5	0.267272755	FALSE
249	2216.0	Children	7.57	5.83	6.41	5	0.268806685	FALSE
250	4614.0	Romance	7.15	4.67	6.62	5	0.269744894	FALSE
251	2030.0	Woman	6.71	4.54	5.6	5	0.269798539	FALSE
252	7260.0	Torte	7.21	5.11	6.03	5	0.27477154	FALSE
253	2391.0	Boy	7.11	4.63	6.11	5	0.276389998	FALSE
254	8496.0	WaterSlide	7.58	5.79	6.33	5	0.276626494	FALSE
255	7405.0	Cupcakes	7.38	6.28	5.67	5	0.28272034	FALSE
256	2345.0	Children	7.41	5.42	6.51	5	0.2840123	FALSE
257	1313.0	Frog	5.65	4.39	5.91	5	0.28402233	FALSE
258	8461.0	HappyTeens	7.22	4.69	6.36	5	0.294303914	FALSE
259	8325.0	RaceCars	5.63	4.47	5.53	5	0.297126003	FALSE
260	2352.1	Kiss	7.27	5.16	6.04	5	0.300169576	FALSE
261	8490.0	RollerCoaster	7.2	6.68	5.37	5	0.307632313	FALSE
262	1410.0	Ferret	7	4.17	6.05	5	0.307979981	FALSE
263	4220.0	EroticFemale	7.31	6.175	5.615	5	0.313339051	FALSE
264	7470.0	Pancakes	7.08	4.64	5.96	5	0.317812317	FALSE
265	1740.0	Owl	6.91	4.27	5.85	5	0.324523104	FALSE
266	2489.0	Musician	5.66	3.8	5.67	5	0.335605153	FALSE
267	2351.0	NursingBaby	5.49	4.74	5.41	5	0.343898744	FALSE
268	4534.0	MaleDancer	5.7	4.16	6.08	5	0.348346573	FALSE
269	2158.0	Children	7.31	5	6.08	5	0.350104777	FALSE
270	1812.0	Elephants	6.83	3.6	5.91	5	0.354764077	FALSE
271	8210.0	Boat	7.53	5.94	5.82	5	0.362107667	FALSE
272	7510.0	Skyscraper	6.05	4.52	4.96	5	0.36249413	FALSE
273	5201.0	Nature	7.06	3.83	6.73	5	0.364088625	FALSE
274	7500.0	Building	5.33	3.26	5.17	5	0.365802645	FALSE
275	7190.0	Clock	5.55	3.84	5.3	5	0.367381508	FALSE
276	7325.0	Watermelon	7.06	3.55	6.56	5	0.373302871	FALSE
277	2209.0	Bride	7.64	5.59	6.53	5	0.37451667	FALSE
278	8330.0	Winner	6.65	4.06	5.56	5	0.374570967	FALSE
279	7270.0	IceCream	7.53	5.76	5.88	5	0.378041079	FALSE
280	8350.0	TennisPlayer	7.18	5.18	5.78	5	0.380236876	FALSE
281	2331.0	Chef	7.24	4.3	6.37	5	0.387101684	FALSE
282	7430.0	Candy	7.11	4.72	5.86	5	0.388124783	FALSE
283	7280.0	Wines	7.2	4.46	6.1	5	0.389370074	FALSE
284	7502.0	Castle	7.75	5.91	6.64	5	0.389462087	FALSE
285	7480.0	Pasta	7.08	4.55	5.88	5	0.391992859	FALSE
286	4616.0	Romance	6.86	4.43	5.66	5	0.401070259	FALSE
287	2020.0	Adult	5.68	3.34	5.99	5	0.402027098	FALSE
288	5001.0	Sunflower	7.16	3.79	6.49	5	0.4169743	FALSE
289	8300.0	Pilot	7.02	6.14	5.31	5	0.422864079	FALSE
290	8170.0	Sailboat	7.63	6.12	5.72	5	0.425082973	FALSE
291	1630.0	Fawn	7.26	4.45	6.12	5	0.430585188	FALSE
292	1463.0	Kittens	7.45	4.79	6.43	5	0.430962861	FALSE
293	5725.0	Field	7.09	3.55	6.23	5	0.442813791	FALSE
294	8420.0	Tubing	7.76	5.56	6.05	5	0.44904628	FALSE
295	8380.0	Athletes	7.56	5.74	5.8	5	0.451632192	FALSE
296	8497.0	CarnivalRide	7.26	4.19	6.22	5	0.454622282	FALSE
297	2347.0	Children	7.83	5.56	6.54	5	0.456413433	FALSE
298	2160.0	Father	7.58	5.16	6.12	5	0.457021866	FALSE
299	1999.0	Mickey	7.43	4.77	6.64	5	0.465028809	FALSE
300	2310.0	Mother	7.06	4.16	5.89	5	0.468209832	FALSE
301	3550.2	Coach	4.92	5.13	5.38	5	0.476664315	FALSE
302	2382.0	Artist	5.67	3.75	5.97	5	0.479792108	FALSE
303	1604.0	Butterfly	7.11	3.3	6.69	5	0.483203597	FALSE
304	2045.0	Baby	7.87	5.47	6.1	5	0.484841299	FALSE
305	2273.0	Boy	5.41	3.52	5.31	5	0.489321351	FALSE
306	1903.0	Shrimp	5.5	4.25	6.01	5	0.490503617	FALSE
307	1540.0	Cat	7.15	4.54	7.01	5	0.493235752	FALSE
308	7183.0	Checkerboard	5.58	3.78	5.71	5	0.506080704	FALSE

No	ImageCode	Description	Valence	Arousal	Dominance1	Cluster	Uncertainty	Under75thQuantile
309	7077.0	Stove	5.12	4.61	5.6	5	0.516598283	FALSE
310	2220.0	MaleFace	5.03	4.93	5.32	5	0.552986319	FALSE
311	7476.0	Ramen	4.99	4.63	5.45	5	0.658077897	FALSE

Appendix B

Stimulus trios: matched images, words and sounds

No	Trio	Type	Code	Description	Valence	Arousal	Dominance	ValenceSD	ArousalSD	DominanceSD
1	1	sound	291	prowler	3.67	6.35	3.86	1.7	1.76	1.87
2	1	word	767	gossip	3.48	5.74	3.57	2.33	2.38	2.26
3	1	image	1201	spider	3.55	6.36	3.87	1.88	2.11	2.3
4	2	sound	370	courtsport	5.94	4.44	5.83	1.66	1.72	1.78
5	2	word	884	moral	6.2	4.49	5.9	1.85	2.28	2.2
6	2	image	4100	maledancers	6.11	4.39	5.93	1.66	1.75	1.71
7	3	sound	243	couplesneeze	3.86	5.19	4.23	1.7	2.06	1.9
8	3	word	862	maniac	3.76	5.39	4.22	2	2.46	2.07
9	3	image	3280	dentalexam	3.72	5.39	4.06	1.89	2.38	1.99
10	4	sound	722	walking	4.83	4.97	4.66	1.22	1.82	1.49
11	4	word	395	skeptical	4.52	4.91	4.5	1.63	1.92	1.61
12	4	image	1645	wolf	4.99	5.14	4.74	1.64	1.99	1.91
13	5	sound	728	paper1	4.72	4.35	5.4	1.26	2.09	1.6
14	5	word	865	manure	3.1	4.17	4.67	1.74	2.09	1.36
15	5	image	1112	snake	4.71	4.6	5.27	1.7	2.44	2.2
16	6	sound	152	tropical	5.23	5.51	4.78	2.28	2.23	2.1
17	6	word	756	foam	6.07	5.26	5.24	2.03	2.54	1.97
18	6	image	1820	crocodile	5.35	5.67	4.66	2.05	2.09	2.3
19	7	sound	113	cows	5.45	4.88	5.36	1.71	1.95	1.64
20	7	word	577	truck	5.47	4.84	5.33	1.88	2.17	1.83
21	7	image	2351	nursingbaby	5.49	4.74	5.41	2.04	2.05	1.81
22	8	sound	206	shower	6.2	4.4	5.62	1.6	1.82	1.61
23	8	word	950	quality	6.25	4.48	5.64	1.59	2.12	1.59
24	8	image	7352	pizza	6.2	4.58	5.58	2.2	2.45	2.07
25	9	sound	226	laughing	7.78	5.42	6.32	1.37	2.13	1.82
26	9	word	210	honest	7.7	5.32	6.24	1.43	1.92	2.13
27	9	image	1710	puppies	8.34	5.41	6.55	1.12	2.34	1.98
28	10	sound	102	cat	4.63	4.91	5.36	2.17	1.97	1.73
29	10	word	895	nasty	3.58	4.89	5	2.38	2.5	2.17
30	10	image	3550.2	coach	4.92	5.13	5.38	1.62	2.24	2.02
31	11	sound	374	sink	5.6	4.23	5.75	1.35	1.89	1.63
32	11	word	924	paint	5.62	4.1	5.75	1.72	2.36	1.71
33	11	image	2600	beer	5.84	4.16	5.84	1.85	1.74	1.63
34	12	sound	260	babiescry	2.04	6.87	3.46	1.39	2.13	2.31
35	12	word	301	pain	2.13	6.5	3.71	1.81	2.49	2.53
36	12	image	3400	severedhand	2.35	6.91	3.65	1.9	2.22	2.21
37	13	sound	289	gunshot	3.08	6.57	3.55	1.71	1.8	2.07
38	13	word	899	nervous	3.29	6.59	3.56	1.47	2.07	1.73
39	13	image	9622	jet	3.1	6.26	3.66	1.9	1.98	2.31
40	14	sound	425	train	5.09	5.15	4.67	1.42	1.54	1.72
41	14	word	901	news	5.3	5.17	4.6	1.67	2.11	1.88
42	14	image	4631	bikercouple	5.36	5.19	4.87	1.86	2.04	1.61
43	15	sound	625	mayday	3.35	6.94	3.26	2.03	1.77	2.13
44	15	word	601	panic	3.12	7.02	3.2	1.84	2.02	1.67
45	15	image	1931	shark	4	6.8	3.51	2.28	2.02	2.54
46	16	sound	150	seagull	6.95	4.38	5.91	1.64	2.22	1.8
47	16	word	466	useful	7.14	4.26	5.93	1.6	2.47	2.1
48	16	image	7480	pasta	7.08	4.55	5.88	1.62	2.42	1.87
49	17	sound	210	eroticmale1	5.72	6.64	5.39	2.26	1.83	2.21
50	17	word	904	noisy	5.02	6.38	4.93	2.02	1.78	1.76
51	17	image	4651	eroticcouple	6.32	6.34	5.8	2.18	2.05	2.15
52	18	sound	808	bugle	6.32	6.35	5.64	1.76	2.15	1.75
53	18	word	520	nude	6.82	6.41	5.96	1.63	2.09	2.29

No	Trio	Type	Code	Description	Valence	Arousal	Dominance	ValenceSD	ArousalSD	DominanceSD
54	18	image	4693	eroticcouple	6.16	6.57	5.46	1.91	1.9	2.02
55	19	sound	252	malesnore	4.01	4.75	4.33	1.87	2.39	1.99
56	19	word	906	noose	3.76	4.39	4.17	1.64	2.08	1.92
57	19	image	7137	cardamage	4.3	4.81	4.5	1.44	1.95	1.92
58	20	sound	720	brushteeth	4.86	4.18	5.76	1.8	1.79	2.2
59	20	word	913	obnoxious	3.5	4.74	5.39	2.18	2.42	2.2
60	20	image	1945	turtle	4.59	4.42	5.57	1.68	2.03	2.07
61	21	sound	380	jackhammer	3.7	6.33	4.18	1.88	1.73	1.93
62	21	word	915	obsession	4.52	6.41	4.77	2.13	2.13	2.38
63	21	image	1200	spider	3.95	6.03	4.33	2.22	2.38	2.38
64	22	sound	716	slotmachine1	7	6.44	6.54	2.17	1.73	2.03
65	22	word	953	quick	6.64	6.57	6.57	1.61	1.78	1.91
66	22	image	8531	sportcar	7.03	5.41	6.77	1.5	2.15	1.69
67	23	sound	270	whistling	6.1	4.23	5.85	1.83	2.06	1.93
68	23	word	558	field	6.2	4.08	5.84	1.37	2.41	1.94
69	23	image	1850	camels	6.15	4.06	5.94	1.52	2.14	1.91
70	24	sound	724	chewing	5.34	4.91	5.8	1.97	1.74	1.85
71	24	word	781	hard	5.22	5.12	5.59	1.82	2.19	1.63
72	24	image	8232	boxer	5.07	5.1	5.57	1.8	2.21	2.07
73	25	sound	810	beethoven	7.51	4.18	6.07	1.66	2.38	1.92
74	25	word	320	politeness	7.18	3.74	5.74	1.5	2.37	1.7
75	25	image	2314	binoculars	7.55	4	6.17	1.24	2.01	1.78
76	26	sound	104	panting	4.96	5.37	5.06	1.68	1.66	1.82
77	26	word	957	razor	4.81	5.36	4.91	2.16	2.44	1.95
78	26	image	9411	boy	4.63	5.37	4.91	1.58	1.97	2.05
79	27	sound	295	couplesobbing	3.27	5.79	3.94	2.39	1.81	2.1
80	27	word	596	knife	3.62	5.8	4.12	2.18	2	2.18
81	27	image	6020	electricchair	3.41	5.58	4.07	1.98	2.01	2.43
82	28	sound	382	shovel	4.33	4.64	4.95	1.42	1.88	1.73
83	28	word	608	skull	4.27	4.75	4.86	1.83	1.85	1.62
84	28	image	1230	spider	4.35	4.44	5.09	1.685	2.33	2.105
85	29	sound	106	growl1	3.37	6.39	3.54	1.64	1.62	1.84
86	29	word	944	pressure	3.38	6.07	3.45	1.61	2.26	2.07
87	29	image	1110	snake	3.84	5.96	3.84	1.89	2.16	2.31
88	30	sound	809	harp	7.44	3.36	6.29	1.41	1.84	1.87
89	30	word	350	relaxed	7	2.39	5.55	1.77	2.13	1.9
90	30	image	5200	flowers	7.36	3.2	6.21	1.52	2.16	1.88
91	31	sound	377	rain1	5.84	3.93	5.7	1.73	1.87	1.89
92	31	word	557	farm	5.53	3.9	5.59	1.85	1.95	1.81
93	31	image	2489	musician	5.66	3.8	5.67	1.44	1.93	1.78
94	32	sound	610	cowboyindians	5.94	6.48	5.31	2.02	2.11	1.77
95	32	word	961	reunion	6.48	6.34	5.64	2.45	2.35	1.95
96	32	image	4672	eroticcouple	6	6.29	5.38	2.04	2.37	2.28
97	33	sound	706	war	4.16	5.3	4.55	1.68	1.83	1.82
98	33	word	962	revolver	4.02	5.55	4.39	2.44	2.39	2.47
99	33	image	1030	snake	4.3	5.46	4.56	2.35	2.43	2.43
100	34	sound	358	writing	4.52	4.87	5.04	1.34	1.98	1.94
101	34	word	221	hungry	3.58	5.13	4.68	2.01	2.44	2.05
102	34	image	1240	spider	4.22	4.92	4.95	1.94	2.17	2.17
103	35	sound	368	crowd5	5.15	4.75	4.6	1.33	1.84	1.66
104	35	word	96	custom	5.85	4.66	5	1.53	2.12	1.87
105	35	image	1908	jellyfish	5.28	4.88	4.75	1.53	2.15	1.6
106	36	sound	719	dentistdrill	2.89	6.91	2.92	1.67	2.02	2.03
107	36	word	798	hurricane	3.34	6.83	3.07	2.12	2.06	2.18
108	36	image	5971	tornado	3.49	6.65	3.3	1.87	2.02	2.42
109	37	sound	360	rollercoaster	6.94	7.54	4.73	2.25	1.97	2.39
110	37	word	384	sex	8.05	7.36	5.75	1.53	1.91	2.25
111	37	image	8492	rollercoaster	7.21	7.31	4.63	2.26	1.64	2.41
112	38	sound	215	eroticcouple2	6.47	7.32	6.02	2.12	1.81	2.02
113	38	word	530	sexy	8.02	7.36	6.82	1.12	1.91	2.13
114	38	image	4664	eroticcouple	6.61	6.72	5.96	2.23	2.08	2.19
115	39	sound	703	busysignal	2.65	5.68	3.26	1.59	1.89	1.92
116	39	word	386	shamed	2.5	4.88	2.98	1.34	2.27	1.94
117	39	image	9920	caraccident	2.5	5.76	3.09	1.52	1.96	2.13
118	40	sound	120	rooster	5.2	5.41	5.04	2.1	2.13	1.93
119	40	word	966	rough	4.74	5.33	4.81	2	2.04	1.7
120	40	image	4005	eroticfemale	5.43	5.02	5.39	2.08	2	2.11
121	41	sound	151	robin	7.12	4.47	5.73	1.56	2.27	1.92
122	41	word	796	humane	6.89	4.5	5.7	1.7	1.91	1.91
123	41	image	2156	family	7.12	4.34	5.82	1.46	2.11	1.72
124	42	sound	296	womencrying	2.06	6.07	3.24	1.22	1.97	1.96
125	42	word	222	hurt	1.9	5.85	3.33	1.26	2.49	2.22
126	42	image	9325	vomit	1.89	6.01	3.22	1.23	2.54	1.96
127	43	sound	224	kids2	6.11	5.64	5.49	1.9	1.89	1.82
128	43	word	370	salute	5.92	5.31	5.46	1.57	2.23	2.05
129	43	image	4007	attractivefem	6.26	5.63	5.57	1.78	2.26	1.95
130	44	sound	220	boylaugh	7.28	6	5.99	1.91	1.99	1.88
131	44	word	987	song	7.1	6.07	5.85	1.97	2.42	2.12
132	44	image	2208	bride	7.35	5.68	6.21	1.68	2.34	1.74

No	Trio	Type	Code	Description	Valence	Arousal	Dominance	ValenceSD	ArousalSD	DominanceSD
133	45	sound	283	fight3	3.05	6.2	3.85	1.72	1.6	2.05
134	45	word	970	scalding	2.82	5.95	3.82	2.12	2.55	2.3
135	45	image	6830	guns	2.82	6.21	3.67	1.81	2.23	2.5
136	46	sound	726	corkpour	6.82	4.51	6.36	1.6	2.08	1.71
137	46	word	818	intellect	6.82	4.75	6.3	1.96	2.5	1.98
138	46	image	7282	cake	6.72	4.77	6.16	1.48	2.08	1.79
139	47	sound	420	carhorns	2.34	7.08	2.7	1.51	2.06	1.8
140	47	word	604	scared	2.78	6.82	2.94	1.99	2.03	2.19
141	47	image	6370	attack	2.7	6.44	3	1.52	2.19	1.87
142	48	sound	246	heartbeat	4.83	4.65	5.07	1.81	2.49	1.86
143	48	word	974	scissors	5.05	4.47	5.16	0.96	1.76	1.84
144	48	image	9422	battleship	4.95	5.09	4.89	1.72	1.92	2.25
145	49	sound	282	fight2	2.92	7.2	3.92	2.34	1.63	2.31
146	49	word	605	scream	3.88	7.04	4.75	2.07	1.96	2.21
147	49	image	3213	surgery	2.96	6.82	3.92	1.94	2	2.44
148	50	sound	723	radio	4.52	4.42	4.93	1.47	1.92	1.9
149	50	word	765	glass	4.75	4.27	5	1.38	2.07	1.46
150	50	image	2410	boy	4.62	4.13	5	1.72	2.29	2.29
151	51	sound	714	siren2	3.1	6.94	3.56	1.67	1.85	1.73
152	51	word	609	snake	3.31	6.82	3.78	2.2	2.1	2.05
153	51	image	3550.1	planecrash	2.35	6.29	3.47	1.39	1.96	2.1
154	52	sound	279	attack1	1.68	7.95	2.3	1.31	2.22	1.94
155	52	word	612	surgery	2.86	6.35	2.75	2.19	2.32	1.86
156	52	image	9414	execution	2.06	6.49	3.11	1.48	2.26	2.23
157	53	sound	813	wedding	7.2	5.89	5.51	1.86	2.4	1.95
158	53	word	212	hopeful	7.1	5.78	5.41	1.46	2.09	1.92
159	53	image	8300	pilot	7.02	6.14	5.31	1.6	2.21	2.31
160	54	sound	820	funkmusic	6.94	5.87	5.97	1.98	1.92	1.8
161	54	word	507	dancer	7.14	6	6.02	1.56	2.2	1.93
162	54	image	7451	hamburger	6.68	5.84	5.85	2.11	2.03	2.24
163	55	sound	322	typewriter	5.01	4.79	5.35	1.82	2.16	1.98
164	55	word	1004	swamp	5.14	4.86	5.29	2.24	2.36	1.63
165	55	image	2220	maleface	5.03	4.93	5.32	1.39	1.65	1.77
166	56	sound	717	slotmachine2	7.32	6.56	6.39	1.64	2.19	2.3
167	56	word	427	talent	7.56	6.27	6.49	1.25	1.8	1.75
168	56	image	4607	eroticcouple	7.03	6.34	6.14	1.84	2.16	2.2
169	57	sound	725	sodafizz	6.61	4.55	6.3	1.8	2.17	1.95
170	57	word	479	virtue	6.22	4.52	6.13	2.06	2.52	2.09
171	57	image	7320	desserts	6.54	4.44	6.22	1.63	2.12	1.95
172	58	sound	216	eroticcouple3	5.97	6.84	5.31	2.06	1.53	2.05
173	58	word	410	startled	4.5	6.93	4.48	1.67	2.24	1.57
174	58	image	4683	eroticcouple	6.17	6.62	5.39	2.07	1.79	2.04
175	59	sound	230	giggling	7.05	4.84	5.77	1.44	1.86	1.55
176	59	word	786	heal	7.09	4.77	5.79	1.46	2.23	1.8
177	59	image	7430	candy	7.11	4.72	5.86	1.78	2.29	2.02
178	60	sound	255	vomit	2.08	6.59	3.23	1.78	2.08	1.98
179	60	word	430	terrible	1.93	6.27	3.58	1.44	2.44	2.34
180	60	image	9187	injureddog	1.81	6.45	3.17	1.36	2.3	2.11
181	61	sound	170	night	5.31	4.6	4.53	2.12	2.07	1.81
182	61	word	869	medicine	5.67	4.4	4.7	2.06	2.36	1.91
183	61	image	7632	airplane	5.22	4.78	4.4	1.69	2.36	2.09
184	62	sound	114	cattle	5.01	6.04	4.56	1.85	1.81	1.75
185	62	word	1000	storm	4.95	5.71	4.54	2.22	2.34	2.04
186	62	image	1310	leopard	4.6	6	4.37	1.62	1.8	1.97
187	63	sound	816	guitar	6.98	5.23	5.84	1.9	2.08	1.88
188	63	word	943	present	6.95	5.12	5.83	1.85	2.39	1.78
189	63	image	7220	pastry	6.91	5.3	5.8	1.74	2.35	2.24
190	64	sound	278	childabuse	1.57	7.27	3.49	1.43	1.6	2.48
191	64	word	435	thief	2.13	6.89	3.79	1.69	2.13	2.55
192	64	image	3130	mutilation	1.58	6.97	3.46	1.24	2.07	2.07
193	65	sound	826	bagpipes	6.21	5.07	5.61	2.12	2.06	1.88
194	65	word	556	face	6.39	5.04	5.67	1.6	2.18	1.58
195	65	image	8250	motorcyclist	6.19	5.04	5.63	1.62	2.49	2.07
196	66	sound	107	dog	5.47	5.85	5.08	2.22	1.81	1.9
197	66	word	1002	sugar	6.74	5.64	5.5	1.73	2.18	1.5
198	66	image	6910	bomber	5.31	5.62	5.1	2.28	2.46	2.46
199	67	sound	171	countrynight	5.59	3.71	5.52	1.79	2.05	1.77
200	67	word	535	golfer	5.61	3.73	5.55	1.93	2.26	1.79
201	67	image	7033	train	5.4	3.99	5.32	1.57	2.14	1.95
202	68	sound	704	phone1	5.49	6.54	5.51	1.98	2.17	1.92
203	68	word	470	vampire	4.26	6.37	5.05	1.86	2.35	2.27
204	68	image	4232	eroticfemale	5.95	6.28	5.69	2.53	2.31	2.52
205	69	sound	261	babycry	2.75	6.51	3.91	1.68	1.96	1.97
206	69	word	471	vandal	2.71	6.4	3.91	1.91	1.88	2.49
207	69	image	6312	abduction	2.48	6.37	3.83	1.52	2.3	2.25
208	70	sound	130	pig	4.64	4.93	5	2.11	1.98	1.91
209	70	word	23	army	4.72	5.03	5.03	1.75	2.03	2.45
210	70	image	9402	mob	4.48	5.07	4.85	2.12	2.15	2.12
211	71	sound	367	casino2	7.33	6.72	6.41	1.74	2.03	1.98

No	Trio	Type	Code	Description	Valence	Arousal	Dominance	ValenceSD	ArousalSD	DominanceSD
212	71	word	819	intercourse	7.36	7	6.4	1.57	2.07	1.78
213	71	image	4608	eroticcouple	7.07	6.47	6.25	1.66	1.96	1.87
214	72	sound	244	manwheeze	2.44	6.31	3.16	1.34	1.85	1.97
215	72	word	616	trauma	2.1	6.33	2.84	1.49	2.45	1.87
216	72	image	9620	shipwreck	2.7	6.11	3.29	1.64	2.1	1.95
217	73	sound	115	bees	2.16	7.03	2.67	1.33	1.91	1.71
218	73	word	618	victim	2.18	6.06	2.69	1.48	2.32	2.04
219	73	image	4664.2	attack	2.79	6.13	3.33	1.77	2.29	2.4
220	74	sound	221	malelaugh	6.56	5.05	5.34	1.75	1.91	1.63
221	74	word	515	gymnast	6.35	5.02	5.31	1.79	2.2	1.79
222	74	image	4150	attractivefem	6.53	4.86	5.45	1.86	2.55	1.81
223	75	sound	611	battletaps	3.02	5.34	3.67	2.06	1.75	1.99
224	75	word	978	severe	3.2	5.26	3.83	1.74	2.36	1.91
225	75	image	7135	cardamage	3.17	5.36	3.76	1.57	2.14	2.19
226	76	sound	112	kids1	6.84	4.46	6.07	1.72	2.13	1.68
227	76	word	761	garden	6.71	4.39	6.02	1.74	2.35	1.71
228	76	image	7390	icecream	6.84	4.56	6.02	1.73	2.28	2.03
229	77	sound	424	carwreck	2.04	7.99	2.29	1.52	1.66	1.74
230	77	word	619	volcano	4.84	6.33	3.25	2.14	2.21	1.97
231	77	image	1050	snake	3.46	6.87	3.08	2.15	1.68	1.93
232	78	sound	910	electricity	3.86	6.18	4.03	1.83	2.27	1.84
233	78	word	484	wasp	3.37	5.5	3.76	1.63	2.17	1.82
234	78	image	6211	attack	3.62	5.9	4.03	2.07	2.22	2.22
235	79	sound	601	colonialmusic	6.53	5.84	5.73	1.66	1.8	1.58
236	79	word	234	interest	6.97	5.66	5.88	1.53	2.26	1.78
237	79	image	2389	teens	6.61	5.63	5.9	1.69	2	1.99
238	80	sound	351	applause1	7.32	5.55	6.74	1.62	2.08	1.71
239	80	word	933	penthouse	6.81	5.52	6.52	1.64	2.49	1.82
240	80	image	4599	romance	7.12	5.69	6.49	1.48	1.94	1.79
241	81	sound	311	crowd2	7.65	7.12	6.09	1.58	1.83	2.18
242	81	word	422	surprised	7.47	7.47	6.11	1.56	2.09	2.19
243	81	image	4670	eroticcouple	6.99	6.74	5.85	1.73	2.03	2.47
244	82	sound	700	toilet	4.68	4.03	5.62	1.61	2.36	1.92
245	82	word	806	immature	3.39	4.15	4.85	1.7	1.96	2.2
246	82	image	7018	screw	4.81	3.91	5.71	0.88	1.97	1.62
247	83	sound	293	mansobbing	3.08	5.74	3.94	1.92	1.69	1.82
248	83	word	712	damage	3.05	5.57	3.88	1.65	2.26	1.86
249	83	image	6220	boysw/guns	3.1	5.89	3.92	1.91	2.43	2.2
250	84	sound	352	sportscrowd	7.17	7.07	5.77	1.97	2.12	2.08
251	84	word	155	exercise	7.13	6.84	5.68	1.58	2.06	2.44
252	84	image	4659	eroticcouple	6.87	6.93	5.67	1.99	2.07	2.52
253	85	sound	375	polaroid	5.99	4.48	5.67	1.6	1.74	1.95
254	85	word	691	coast	5.98	4.59	5.67	1.86	2.31	1.71
255	85	image	8325	racecars	5.63	4.47	5.53	1.5	2.19	1.82
256	86	sound	109	carousel	6.4	5.64	5.69	2.13	1.84	1.93
257	86	word	639	answer	6.63	5.41	5.85	1.68	2.43	1.88
258	86	image	8021	skier	6.79	5.67	5.85	1.44	2.37	2.06
259	87	sound	204	eroticfem4	5.68	6.82	5.34	2.16	1.71	2.13
260	87	word	21	anxious	4.81	6.92	5.33	1.98	1.81	1.82
261	87	image	4698	eroticcouple	6.5	6.72	5.7	1.67	1.72	2.1
262	88	sound	699	bomb	3.59	6.15	3.47	2.07	2.36	1.97
263	88	word	207	hide	4.32	5.28	3.4	1.91	2.51	2.12
264	88	image	5972	tornado	3.85	6.34	3.49	2.33	2.2	2.42
265	89	sound	363	horserace	6.1	6.32	5.05	1.88	2	1.67
266	89	word	28	astonished	6.56	6.58	5.16	1.61	2.22	1.79
267	89	image	8206	surfers	6.43	6.41	5.19	1.75	2.19	2.04
268	90	sound	802	nativesong	6.17	5.29	5.72	1.99	1.74	1.8
269	90	word	570	red	6.41	5.29	5.78	1.61	2.04	1.59
270	90	image	8280	diver	6.38	5.05	5.85	1.46	2.18	1.95
271	91	sound	205	eroticfem3	6.47	6.46	5.81	1.98	2.06	1.94
272	91	word	644	athletics	6.61	6.1	6.12	2.08	2.29	2.12
273	91	image	4658	eroticcouple	6.62	6.47	5.86	1.89	2.14	2.35
274	92	sound	817	bongos	7.67	7.15	6.44	1.46	2.11	1.73
275	92	word	630	adventure	7.6	6.98	6.46	1.5	2.15	1.67
276	92	image	4687	eroticcouple	6.87	6.51	6.04	1.51	2.1	1.96
277	93	sound	319	office2	3.56	6.08	3.82	1.44	1.39	1.79
278	93	word	645	avalanche	3.29	5.54	3.61	1.95	2.37	2
279	93	image	1113	snake	3.81	6.06	3.91	1.75	2.12	2.1
280	94	sound	250	malesneeze	3.54	4.94	4.08	1.57	1.9	1.67
281	94	word	842	lawsuit	3.37	4.93	3.92	2	2.44	2.02
282	94	image	7136	carboot	3.47	5.01	3.98	1.7	2.17	2.23
283	95	sound	200	eroticcouple	6.31	7.1	5.92	1.93	1.66	2
284	95	word	11	alert	6.2	6.85	5.96	1.76	2.53	2.24
285	95	image	4656	eroticcouple	6.73	6.41	6.1	1.94	2.19	2.05
286	96	sound	172	brook	6.62	3.36	6.21	1.69	2.07	1.86
287	96	word	547	basket	5.45	3.63	5.76	1.15	2.02	1.45
288	96	image	2560	picnic	6.34	3.49	6.06	1.53	2.07	1.85
289	97	sound	730	glassbreak	3.22	6.23	4.1	1.45	1.78	1.87
290	97	word	33	bastard	3.36	6.07	4.17	2.16	2.15	2.4

No	Trio	Type	Code	Description	Valence	Arousal	Dominance	ValenceSD	ArousalSD	DominanceSD
291	97	image	9321	vomit	2.81	6.24	3.9	2.14	2.23	2.35
292	98	sound	698	rain2	5.18	4.12	4.85	1.94	1.98	1.96
293	98	word	283	month	5.15	4.03	4.85	1.09	1.77	1.14
294	98	image	7190	clock	5.55	3.84	5.3	1.34	2.06	2.04
295	99	sound	132	chickens	5.64	4.77	5.96	1.76	1.73	1.82
296	99	word	824	invest	5.93	5.12	5.88	2.1	2.42	1.95
297	99	image	1121	lizard	5.79	4.83	5.89	1.61	1.98	1.99
298	100	sound	292	malescream	1.99	7.28	2.82	1.41	1.74	1.78
299	100	word	15	ambulance	2.47	7.33	3.22	1.5	1.96	2.29
300	100	image	9412	deadman	1.83	6.72	3	1.37	2.07	2.32
301	101	sound	110	baby	7.64	6.03	6.14	2.1	1.98	1.88
302	101	word	514	food	7.65	5.92	6.18	1.37	2.11	2.48
303	101	image	8470	gymnast	7.74	6.14	6.17	1.53	2.19	2.09
304	102	sound	320	office1	4.23	5.48	4.81	1.56	1.95	1.85
305	102	word	141	embattled	4.39	5.36	4.81	1.63	2.37	1.79
306	102	image	2458	cryingbaby	4.69	5.28	5.06	1.88	1.88	1.84
307	103	sound	361	restaurant	5.36	5.01	5.25	1.62	1.65	1.6
308	103	word	543	black	5.39	4.61	5.14	1.8	2.24	1.79
309	103	image	4460	eroticmale	5.6	4.94	5.34	1.61	2.09	1.79
310	104	sound	721	beer	6.71	5	5.96	1.75	2.12	1.71
311	104	word	934	perfume	6.76	5.05	5.93	1.48	2.36	1.69
312	104	image	7450	cheeseburger	6.4	5.05	5.81	2.01	2.22	2.2
313	105	sound	364	bar	5.19	5.62	4.83	1.85	1.75	1.67
314	105	word	585	boxer	5.51	5.12	5.1	1.8	2.26	1.64
315	105	image	7600	dragon	5.9	5.5	5.22	1.76	1.92	2.01
316	106	sound	312	crowd3	3.89	6.89	3.68	2.13	1.88	1.94
317	106	word	684	chaos	4.17	6.67	3.86	2.36	2.06	1.95
318	106	image	1040	snake	3.99	6.25	3.92	2.24	2.13	2.13
319	107	sound	288	creep	2.71	6.82	3.59	1.75	1.63	2.21
320	107	word	89	crash	2.31	6.95	3.44	1.44	2.44	2.21
321	107	image	9250	warvictim	2.57	6.6	3.73	1.39	1.87	1.94
322	108	sound	710	cuckoo	4.27	6.24	4.2	2.04	1.88	1.77
323	108	word	553	cliff	4.67	6.25	4.35	2.08	2.15	2.11
324	108	image	1101	snake	4.1	5.83	4.13	1.85	2.25	2.3
325	109	sound	708	clock	4.34	3.51	4.64	1.42	2.05	2.06
326	109	word	405	solemn	4.32	3.56	4.61	1.51	1.95	1.87
327	109	image	2206	fingerprint	4.06	3.71	4.46	1.4	2.03	2.36
328	110	sound	415	countdown	6.46	6.55	4.8	1.67	1.56	2.25
329	110	word	680	casino	6.81	6.51	5.12	1.66	2.12	2.15
330	110	image	8180	cliffdivers	7.12	6.59	4.97	1.88	2.12	2.52
331	111	sound	500	wind	4.32	5.4	4.24	2.03	1.93	1.75
332	111	word	689	clumsy	4	5.18	3.86	2.22	2.4	1.79
333	111	image	3210	surgery	4.49	5.39	4.3	1.91	1.91	2.18
334	112	sound	116	buzzing	3.02	6.51	4.14	1.65	2.13	2.11
335	112	word	583	bees	3.2	6.51	4.16	2.07	2.14	2.11
336	112	image	3212	surgery	2.79	6.57	4.07	1.67	1.99	2.25
337	113	sound	111	musicbox	6.01	5.65	5.42	2.19	1.91	2.02
338	113	word	95	curious	6.08	5.82	5.42	1.63	1.64	1.6
339	113	image	4008	eroticfemale	5.91	5.66	5.34	2.24	2.32	2.07
340	114	sound	373	paint	5.09	4.65	5.69	1.55	2.17	1.78
341	114	word	834	kick	4.31	4.9	5.5	2.18	2.35	1.93
342	114	image	4004	eroticfemale	5.14	4.44	5.75	1.85	2.14	2.13
343	115	sound	624	airraid	2.82	7.1	3.41	1.75	2.1	2.03
344	115	word	713	danger	2.95	7.32	3.59	2.22	2.07	2.31
345	115	image	3150	mutilation	2.26	6.55	3.39	1.57	2.2	2.15
346	116	sound	376	lawnmower	4.88	4.6	5.19	1.62	1.93	1.63
347	116	word	84	context	5.2	4.22	5.17	1.38	2.24	1.39
348	116	image	2309	girlcow	4.89	4.33	5.39	1.71	1.92	1.64
349	117	sound	400	jet	6.02	5.38	4.86	1.49	1.87	1.86
350	117	word	633	alien	5.6	5.45	4.64	1.82	2.15	2.07
351	117	image	1640	coyote	6.215	5.155	5.065	2.05	2.065	2.1
352	118	sound	353	baseball	7.38	6.62	6.04	1.53	1.42	1.86
353	118	word	506	couple	7.41	6.39	6.02	1.97	2.31	2.28
354	118	image	8200	waterskier	7.54	6.35	6.17	1.37	1.98	1.61
355	119	sound	133	growl2	3.79	6.23	3.61	1.69	1.84	1.8
356	119	word	589	dentist	4.02	5.73	3.8	2.23	2.13	2.16
357	119	image	1070	snake	3.96	6.16	3.71	2.3	2.08	2.08
358	120	sound	284	attack3	2.01	7.05	2.99	1.48	1.65	2
359	120	word	591	drown	1.92	6.57	2.86	1.48	2.33	1.99
360	120	image	3110	burnvictim	1.79	6.7	3.04	1.3	2.16	1.97
361	121	sound	280	womancrying	3.65	5.33	4.37	1.87	1.46	1.95
362	121	word	436	thorn	3.64	5.14	4.45	1.76	2.14	1.5
363	121	image	9594	injection	3.76	5.17	4.43	1.7	2.17	2.25
364	122	sound	241	malecough	2.46	5.87	3.52	1.53	2.06	2.07
365	122	word	348	regretful	2.28	5.74	3.43	1.42	2.32	2.52
366	122	image	9611	planecrash	2.71	5.75	3.67	1.95	2.44	2.23
367	123	sound	378	doorbell	6.06	6.15	5.47	2.01	2.22	1.83
368	123	word	137	education	6.69	5.74	6.15	1.77	2.46	2.35
369	123	image	4666	eroticcouple	6.24	6.1	5.55	1.78	2.2	2.1

No	Trio	Type	Code	Description	Valence	Arousal	Dominance	ValenceSD	ArousalSD	DominanceSD
370	124	sound	811	bach	7.4	4.95	6.14	1.63	2.46	1.87
371	124	word	116	devoted	7.41	5.23	6.18	1.37	2.21	2.36
372	124	image	2158	children	7.31	5	6.08	1.48	2.2	1.62
373	125	sound	701	fan	4.95	4.41	5.27	1.47	2.06	1.62
374	125	word	1001	stove	4.98	4.51	5.36	1.69	2.14	1.87
375	125	image	7182	checkerboard	5.16	4.02	5.51	1.31	2.12	2.1
376	126	sound	626	explosion	3.37	6.61	3.4	1.98	1.71	1.86
377	126	word	140	embarrassed	3.03	5.87	2.87	1.85	2.55	1.99
378	126	image	1052	snake	3.5	6.52	3.36	1.87	2.23	2.26
379	127	sound	202	eroticfem2	6.81	7.13	6.16	2.08	1.89	2.18
380	127	word	512	erotic	7.43	7.24	6.39	1.53	1.97	2.16
381	127	image	4652	eroticcouple	6.79	6.62	6.1	2.02	2.04	2.22
382	128	sound	201	eroticfem1	6.7	7.31	5.93	2.22	1.86	2.4
383	128	word	152	excitement	7.5	7.67	6.18	2.2	1.91	2.17
384	128	image	4668	eroticcouple	6.67	7.13	5.73	1.69	1.62	2.34
385	129	sound	423	injury	3.31	6.23	4.22	1.79	1.6	1.89
386	129	word	80	confused	3.21	6.03	4.24	1.51	1.88	1.91
387	129	image	3250	openchest	3.78	6.29	4.45	1.72	1.63	1.99
388	130	sound	245	hiccup	4.18	5.05	4.26	1.85	1.82	1.81
389	130	word	743	fall	4.09	4.7	4	2.21	2.48	2.15
390	130	image	9270	toxicwaste	3.72	5.24	4.04	1.51	2.15	2.05
391	131	sound	815	rocknroll	7.9	6.85	6.86	1.53	2.16	1.99
392	131	word	157	fame	7.93	6.55	6.85	1.29	2.46	2.14
393	131	image	8034	skier	7.06	6.3	6.26	1.53	2.16	2.02
394	132	sound	355	crowd4	6.77	6.32	5.7	1.84	1.66	2
395	132	word	760	game	6.98	5.89	5.7	1.97	2.37	1.65
396	132	image	7650	city	6.62	6.15	5.79	1.91	2.24	1.98
397	133	sound	732	crash	2.89	6.98	3.32	1.68	1.75	1.88
398	133	word	592	fear	2.76	6.96	3.22	2.12	2.17	2.2
399	133	image	6821	gang	2.38	6.29	3.29	1.72	2.02	2.36
400	134	sound	254	videogame	6.17	5.58	6.25	1.65	1.99	2.05
401	134	word	1005	swift	6.46	5.39	6.29	1.76	2.53	1.85
402	134	image	7240	gym	6.02	5.51	6.37	1.93	2.12	2.42
403	135	sound	105	puppy	2.88	6.4	3.8	2.14	2.13	2.17
404	135	word	163	fearful	2.25	6.33	3.64	1.18	2.28	2.18
405	135	image	6834	police	2.91	6.28	3.9	1.73	1.9	2.28
406	136	sound	365	party	6.97	6.32	5.73	1.9	1.9	1.76
407	136	word	749	festive	7.3	6.58	5.77	2.26	2.29	2.34
408	136	image	5629	hiker	7.03	6.55	5.68	1.55	2.11	2.55
409	137	sound	225	clapgame	5.96	4.83	5.49	1.51	1.93	1.56
410	137	word	753	flag	6.02	4.6	5.5	1.66	2.35	1.66
411	137	image	4470	eroticmale	5.87	4.81	5.45	1.63	2.31	1.61
412	138	sound	709	alarmclock	2.78	7.54	3.95	1.93	2.28	2.24
413	138	word	166	fire	3.22	7.17	4.49	2.06	2.06	2.49
414	138	image	2730	nativeboy	2.45	6.8	3.94	2.25	2.21	2.55
415	139	sound	627	rain1	4.83	4.65	4.53	1.89	2.12	1.65
416	139	word	434	theory	5.3	4.62	4.88	1.49	1.94	1.81
417	139	image	9468	grafitti	4.67	4.68	4.58	1.8	1.89	2.09
418	140	sound	502	enginefailure	3.15	6.32	3.23	2.01	1.87	2.1
419	140	word	755	flood	3.19	6	3.24	1.66	2.02	2.14
420	140	image	1304	attackdog	3.37	6.37	3.29	1.58	1.93	1.67
421	141	sound	602	thunderstorm	5.99	3.77	4.85	2.23	1.74	2.27
422	141	word	219	humble	5.86	3.74	4.76	1.42	2.33	2.25
423	141	image	5991	sky	6.55	4.01	4.78	2.09	2.44	2.17
424	142	sound	729	paper2	4.3	5.79	5.33	1.69	1.9	2.27
425	142	word	9	aggressive	5.1	5.83	5.59	1.68	2.33	2.4
426	142	image	8060	boxer	5.36	5.31	5.92	2.23	1.99	2.43
427	143	sound	366	casino1	7.09	6.26	6.08	1.73	1.63	2.19
428	143	word	654	beautiful	7.6	6.17	6.29	1.64	2.34	1.81
429	143	image	4680	eroticcouple	7.25	6.02	6.27	1.83	2.27	2.29
430	144	sound	705	phone2	5.35	4.15	5.68	1.43	1.72	1.81
431	144	word	568	office	5.24	4.08	5.59	1.59	1.92	1.89
432	144	image	2372	woman	5.48	4.09	5.72	1.63	1.99	2.01
433	145	sound	262	yawn	5.26	2.88	4.87	1.58	1.74	1.83
434	145	word	309	pencil	5.22	3.14	4.78	0.68	1.9	1.73
435	145	image	9210	rain	4.53	3.08	4.55	1.82	2.13	1.9
436	146	sound	403	helicopter1	5.57	5.56	5.31	1.83	1.99	1.96
437	146	word	878	mischief	5.57	5.76	5.56	2.05	1.95	1.88
438	146	image	4325	attractivefem	5.96	5.18	5.58	1.65	2.19	1.78
439	147	sound	410	helicopter2	4.86	5.89	4.59	1.48	2.06	1.55
440	147	word	215	hospital	5.04	5.98	4.69	2.45	2.54	2.16
441	147	image	1080	snake	4.24	5.69	4.33	2.08	2.28	2.28
442	148	sound	242	femalecough	2.8	5.39	3.76	1.86	1.91	1.81
443	148	word	917	offend	2.76	5.56	3.73	1.5	2.06	2.03
444	148	image	9610	accident	2.89	5.23	3.82	1.43	2.14	2.05
445	149	sound	251	noseblow	4.16	5.14	4.44	2.02	2.11	1.89
446	149	word	914	obscene	4.23	5.04	4.48	2.3	2.3	1.91
447	149	image	9582	dentalexam	4.18	5.29	4.33	2.28	2.21	2.29

Appendix C

Stimulus quartets: matched images, words and sounds, with texts

No	Pair	Type	Status	Code	Descr	Valence	Arousal	Dominance	Cluster
1	1	image	ActualStimulus	3130	mutilation	1.58	6.97	3.46	1
2	1	text	ExplanatoryText	9500	You are leaving the concert when a drunk, smelling of smoke and alcohol, stumbles into you and throws up on your jacket. You retch as vomit drips onto your hand.	1.7	7.09	3.42	1
3	2	image	ActualStimulus	9183	hurtdog	1.69	6.58	2.96	1
4	2	text	ExplanatoryText	7340	Before smelling the rotten meat, you take a huge bite of the hamburger.	2.54	6.62	3.22	1
5	3	image	ActualStimulus	9412	deadman	1.83	6.72	3	1
6	3	text	ExplanatoryText	9180	A rabbit darts in front of your car, you are unable to avoid hitting it.	2.54	6.9	3.52	1
7	4	image	ActualStimulus	9187	injureddog	1.81	6.45	3.17	1
8	4	text	ExplanatoryText	9320	You gag as you enter the filthy bathroom, the toilet has overflowed.	2.38	5.85	3.93	1
9	5	image	ActualStimulus	3110	burnvictim	1.79	6.7	3.04	1
10	5	text	ExplanatoryText	1350	You jump back, muscles tense, as the large dog strains against the chain, slobbering with teeth bared, leaping and snarling in a crazy rage.	2.98	7.1	3.91	1
11	6	sound	ActualStimulus	115	bees	2.16	7.03	2.67	1
12	6	text	ExplanatoryText	2600	You've waited endlessly at the check-out counter. Trapped. Others crowd against you. There's a sudden rushing in your head. You gasp for breath, chest tight, temples throbbing. Is it a heart attack?	2.07	7.48	2.51	1
13	7	sound	ActualStimulus	420	carhorns	2.34	7.08	2.7	1
14	7	text	ExplanatoryText	8020	It's your turn to speak to the group. They're all looking at you. Your mouth's dry and you can't get the words out. Your heart pounds in the silent room. Someone laughs.	2.62	7.17	2.84	1
15	8	sound	ActualStimulus	292	malescream	1.99	7.28	2.82	1
16	8	text	ExplanatoryText	9190	Your pet is lost, you gasp when you see a smear of blood on the road.	1.87	7.32	2.75	1
17	9	sound	ActualStimulus	424	carwreck	2.04	7.99	2.29	1
18	9	text	ExplanatoryText	1500	The snake darts forward, jaws open, and sinks its fangs into your leg.	1.89	7.94	2.27	1
19	10	sound	ActualStimulus	279	attack1	1.68	7.95	2.3	1
20	10	text	ExplanatoryText	1960	You step on something in the leaves, and suddenly a snake ' mouth gaping, fangs protruding'is hurtling at you. It strikes, you feel a sharp, stabbing pain.	1.7	7.65	2.32	1
21	11	word	ActualStimulus	604	scared	2.78	6.82	2.94	1
22	11	text	ExplanatoryText	5800	As you ease the car onto the wooden bridge, it groans. In the headlights, a broken railing swings in the wind. A swift current rams against the pilings below.	2.86	6.81	3	1
23	2	word	ActualStimulus	591	drown	1.92	6.57	2.86	1
24	2	text	ExplanatoryText	7340	Before smelling the rotten meat, you take a huge bite of the hamburger.	2.54	6.62	3.22	1

No	Pair	Type	Status	Code	Descr	Valence	Arousal	Dominance	Cluster
25	12	word	ActualStimulus	616	trauma	2.1	6.33	2.84	1
26	12	text	ExplanatoryText	8030	Everyone's talking, laughing together at the party. You're alone– tense, sweaty. People glance at you and quickly look away. When asked your name, throat dry, you croak an answer.	2.66	6.2	2.93	1
27	13	word	ActualStimulus	618	victim	2.18	6.06	2.69	1
28	13	text	ExplanatoryText	9580	Above his mask, the dentist frowns in concentration. He presses a hooked probe into your gum. Saliva flows and you start to gag. The technician adjusts the suction pipe.	2.81	6.1	3.04	1
29	14	word	ActualStimulus	15	ambulance	2.47	7.33	3.22	1
30	14	text	ExplanatoryText	6800	It's late at night in a poorly lit parking lot. You tense, clutching the keys. Your car stands alone in the distance, when footsteps sound behind you.	2.5	7.5	3.3	1
31	15	image	ActualStimulus	5200	flowers	7.36	3.2	6.21	2
32	15	text	ExplanatoryText	5620	You lie on the warm sand listening to the sound of screeching gulls.	7.43	3.17	6	2
33	15	image	ActualStimulus	2314	binoculars	7.55	4	6.17	2
34	15	text	ExplanatoryText	5620	You lie on the warm sand listening to the sound of screeching gulls.	7.43	3.17	6	2
35	16	image	ActualStimulus	7480	pasta	7.08	4.55	5.88	2
36	16	text	ExplanatoryText	7355	When the pizza arrives, you sink your teeth into thick layers of cheese.	7.65	5.52	6.24	2
37	17	image	ActualStimulus	7390	icecream	6.84	4.56	6.02	2
38	17	text	ExplanatoryText	8660	You won a free pass to the whole carnival. Like kids again, you all jump on the merry-go-round, laughing as it turns, singing along with the music: 'What a wonderful day!'	7.75	5.72	6.98	2
39	18	image	ActualStimulus	2156	family	7.12	4.34	5.82	2
40	18	text	ExplanatoryText	1620	Hiking around the national park, you see a fawn nuzzling her mother.	7.16	4.63	5.31	2
41	19	sound	ActualStimulus	810	beethoven	7.51	4.18	6.07	2
42	19	text	ExplanatoryText	2560	You lounge around the crowded table, laughing with your family.	8	5.05	5.95	2
43	20	sound	ActualStimulus	112	kids1	6.84	4.46	6.07	2
44	20	text	ExplanatoryText	5500	The mountain air is clear and cold. The sun glistens on the powder as you head down the slope in gliding turns, mastering the mountain, moving with a sure, easy grace.	7.71	5.16	7.35	2
45	15	sound	ActualStimulus	809	harp	7.44	3.36	6.29	2
46	15	text	ExplanatoryText	5620	You lie on the warm sand listening to the sound of screeching gulls.	7.43	3.17	6	2
47	21	sound	ActualStimulus	150	seagull	6.95	4.38	5.91	2
48	21	text	ExplanatoryText	1460	Your new kitten nestles comfortably in your lap as you stroke her fur.	7.44	3.53	6.69	2
49	18	sound	ActualStimulus	151	robin	7.12	4.47	5.73	2
50	18	text	ExplanatoryText	1620	Hiking around the national park, you see a fawn nuzzling her mother.	7.16	4.63	5.31	2
51	19	word	ActualStimulus	786	heal	7.09	4.77	5.79	2
52	19	text	ExplanatoryText	2560	You lounge around the crowded table, laughing with your family.	8	5.05	5.95	2
53	19	word	ActualStimulus	796	humane	6.89	4.5	5.7	2
54	19	text	ExplanatoryText	2560	You lounge around the crowded table, laughing with your family.	8	5.05	5.95	2
55	15	word	ActualStimulus	320	politeness	7.18	3.74	5.74	2
56	15	text	ExplanatoryText	5620	You lie on the warm sand listening to the sound of screeching gulls.	7.43	3.17	6	2
57	22	word	ActualStimulus	761	garden	6.71	4.39	6.02	2
58	22	text	ExplanatoryText	2840	The tangles easily come out of your hair as you carefully brush through it.	6.13	3.51	6.22	2
59	21	word	ActualStimulus	466	useful	7.14	4.26	5.93	2
60	21	text	ExplanatoryText	1460	Your new kitten nestles comfortably in your lap as you stroke her fur.	7.44	3.53	6.69	2
61	23	image	ActualStimulus	9210	rain	4.53	3.08	4.55	3
62	23	text	ExplanatoryText	2530	You've been sick all week, lying on a lumpy couch with a bad cold.	2.15	3.32	3.09	3
63	24	image	ActualStimulus	2206	fingerprint	4.06	3.71	4.46	3
64	24	text	ExplanatoryText	2610	You are sitting at the kitchen table with yesterday's newspaper in front of you. You push back the chair when you hear the coffee maker slow to a stop.	5.37	3.13	6.11	3
65	25	image	ActualStimulus	1230	spider	4.35	4.44	5.09	3
66	25	text	ExplanatoryText	7040	You hold the flashlight steady in order to get a better look at the map.	5.04	4.17	5.65	3

No	Pair	Type	Status	Code	Descr	Valence	Arousal	Dominance	Cluster
67	26	image	ActualStimulus	1240	spider	4.22	4.92	4.95	3
68	26	text	ExplanatoryText	9450	Something is caught on the sole of your shoe. You reach down, and your hand comes away sticky with someone's gum. A piece adheres to your fingers.	2.64	5.3	4.06	3
69	27	image	ActualStimulus	2410	boy	4.62	4.13	5	3
70	27	text	ExplanatoryText	2120	You wash and rinse the dishes in the sink while your roommate vacuums.	5.36	4.06	5.68	3
71	28	sound	ActualStimulus	700	toilet	4.68	4.03	5.62	3
72	28	text	ExplanatoryText	2580	You run the comb through your hair, straighten your collar, smooth out the shirt's wrinkles. Water is running in the sink. You turn it off and leave.	5.55	3.6	6.46	3
73	29	sound	ActualStimulus	708	clock	4.34	3.51	4.64	3
74	29	text	ExplanatoryText	2590	Sitting on the couch with the remote, you aimlessly flip through TV channels.	5.07	2.89	5.78	3
75	30	sound	ActualStimulus	358	writing	4.52	4.87	5.04	3
76	30	text	ExplanatoryText	7595	You walk through the crowded parking lot, heading for your car.	5	4.47	5.56	3
77	30	sound	ActualStimulus	382	shovel	4.33	4.64	4.95	3
78	30	text	ExplanatoryText	7595	You walk through the crowded parking lot, heading for your car.	5	4.47	5.56	3
79	27	sound	ActualStimulus	723	radio	4.52	4.42	4.93	3
80	27	text	ExplanatoryText	2120	You wash and rinse the dishes in the sink while your roommate vacuums.	5.36	4.06	5.68	3
81	31	word	ActualStimulus	806	immature	3.39	4.15	4.85	3
82	31	text	ExplanatoryText	2540	You walk through the supermarket aisles checking things off your list as you pick each item you need off the shelves.	5.54	3.38	6.7	3
83	30	word	ActualStimulus	608	skull	4.27	4.75	4.86	3
84	30	text	ExplanatoryText	7595	You walk through the crowded parking lot, heading for your car.	5	4.47	5.56	3
85	32	word	ActualStimulus	913	obnoxious	3.5	4.74	5.39	3
86	32	text	ExplanatoryText	8780	Doing an overhead press, you exhale with force as you push the bar up. You hold the weight at the top for a moment and then slowly bring it down, keeping good form.	6.12	5.76	7.16	3
87	33	word	ActualStimulus	834	kick	4.31	4.9	5.5	3
88	33	text	ExplanatoryText	1950	You watch a giant snake coiled in a display case. You freeze, as the snake's eyes move in your direction, and a red forked-tongue darts out.	4.2	5.81	5.27	3
89	34	word	ActualStimulus	405	solemn	4.32	3.56	4.61	3
90	34	text	ExplanatoryText	2230	You sit at the kitchen table, drinking your morning coffee and reading the paper.	6.08	3.43	6.37	3
91	35	image	ActualStimulus	4668	eroticcouple	6.67	7.13	5.73	4
92	35	text	ExplanatoryText	8025	Your heart pounds as you begin your speech in the auditorium.	4.92	7.89	4.4	4
93	36	image	ActualStimulus	4659	eroticcouple	6.87	6.93	5.67	4
94	36	text	ExplanatoryText	8033	The band is terrific. The room vibrates with sound and your skin tingles. You're dancing together, moving effortlessly with the music. You're feeling great!	8.22	6.62	6.72	4
95	37	image	ActualStimulus	4670	eroticcouple	6.99	6.74	5.85	4
96	37	text	ExplanatoryText	8190	Skiing down the mountain slopes, you glide easily over the powdery snow.	7.51	6.17	6.63	4
97	38	image	ActualStimulus	8492	rollercoaster	7.21	7.31	4.63	4
98	38	text	ExplanatoryText	8480	You tense as the roller coaster reaches the crest. Then, you are all plunging down, screaming above the roar, together, laughing, and waving your arms.	7.58	7.95	4.66	4
99	39	image	ActualStimulus	8200	waterskier	7.54	6.35	6.17	4
100	39	text	ExplanatoryText	4100	You raise your champagne glass and greet the new year with your lover.	8.22	6.65	6.44	4
101	40	sound	ActualStimulus	311	crowd2	7.65	7.12	6.09	4
102	40	text	ExplanatoryText	7496	You dance in the packed bar as your favorite DJ spins the tunes.	7.98	6.97	6.43	4
103	41	sound	ActualStimulus	817	bongos	7.67	7.15	6.44	4
104	41	text	ExplanatoryText	2770	She really likes your gift. As soon she saw it, she screamed with joy: 'Thank you. It's just perfect. Fantastic!' Your heart beats with pleasure, when she leaps up and hugs you.	8.2	6.49	6.89	4
105	42	sound	ActualStimulus	201	eroticfem1	6.7	7.31	5.93	4

No	Pair	Type	Status	Code	Descr	Valence	Arousal	Dominance	Cluster
106	42	text	ExplanatoryText	8790	Working hard and feeling the burn, you try to finish your last set of crunches. As you complete the last repetition, your pace is very slow and your abs quiver in exhaustion.	6.22	6.24	6.87	4
107	43	sound	ActualStimulus	360	rollercoaster	6.94	7.54	4.73	4
108	43	text	ExplanatoryText	4300	As soon as you saw each other, the affair began. You remember beautiful eyes looking straight into yours– your heart in your throat, at the first touch.	7.59	7.63	5.23	4
109	44	sound	ActualStimulus	815	rocknroll	7.9	6.85	6.86	4
110	44	text	ExplanatoryText	2570	The boss smiles and shakes your hand. 'You'll receive a very big raise in pay. Good work!' he says. Your heart skips a beat. Someone shouts congratulations. You smile back.	8.31	6.71	7.41	4
111	45	word	ActualStimulus	384	sex	8.05	7.36	5.75	4
112	45	text	ExplanatoryText	4680	Together in bed, you feel the the gentle touch of naked skin against you.	8.15	7.23	5.82	4
113	46	word	ActualStimulus	152	excitement	7.5	7.67	6.18	4
114	46	text	ExplanatoryText	4650	You are both aroused, breathless. You fall together on the couch. Kisses on your neck, face– warm hands fumbling with clothing, hearts pounding.	8.34	8.1	6.2	4
115	47	word	ActualStimulus	530	sexy	8.02	7.36	6.82	4
116	47	text	ExplanatoryText	8550	All eyes are on you as you walk into the dance with a beautiful date.	8.22	7.17	7.06	4
117	48	word	ActualStimulus	422	surprised	7.47	7.47	6.11	4
118	48	text	ExplanatoryText	8600	It's the last few minutes of the big game and it's close. The crowd explodes in a deafening roar. You jump up, cheering. Your team has come from behind to win.	8.28	7.64	6.48	4
119	49	word	ActualStimulus	512	erotic	7.43	7.24	6.39	4
120	49	text	ExplanatoryText	8760	Running sprints, you breath hard and pump your arms. Your legs feel heavy and your hamstrings burn as your come to the finish line.	6.43	6.94	6.77	4
121	50	image	ActualStimulus	2220	maleface	5.03	4.93	5.32	5
122	50	text	ExplanatoryText	2850	You unfold the map, spread it out on the table, and with your finger trace a route south towards the beach. You refold the map, pick up your bag, and leave.	6.84	4.46	7.06	5
123	51	image	ActualStimulus	1908	jellyfish	5.28	4.88	4.75	5
124	51	text	ExplanatoryText	2880	Your friend whispers to you in a meeting, and you strain to catch the words.	4.63	4.96	4.4	5
125	51	image	ActualStimulus	1645	wolf	4.99	5.14	4.74	5
126	51	text	ExplanatoryText	2880	Your friend whispers to you in a meeting, and you strain to catch the words.	4.63	4.96	4.4	5
127	51	image	ActualStimulus	9422	battleship	4.95	5.09	4.89	5
128	51	text	ExplanatoryText	2880	Your friend whispers to you in a meeting, and you strain to catch the words.	4.63	4.96	4.4	5
129	52	image	ActualStimulus	3550.2	coach	4.92	5.13	5.38	5
130	52	text	ExplanatoryText	8350	Swimming laps, you are on a good pace. You turn your head to inhale, then prepare for your next flip turn. You push off the wall and glide before beginning your stroke again.	6.39	5.18	6.84	5
131	30	sound	ActualStimulus	130	pig	4.64	4.93	5	5
132	30	text	ExplanatoryText	7595	You walk through the crowded parking lot, heading for your car.	5	4.47	5.56	5
133	51	sound	ActualStimulus	722	walking	4.83	4.97	4.66	5
134	51	text	ExplanatoryText	2880	Your friend whispers to you in a meeting, and you strain to catch the words.	4.63	4.96	4.4	5
135	51	sound	ActualStimulus	425	train	5.09	5.15	4.67	5
136	51	text	ExplanatoryText	2880	Your friend whispers to you in a meeting, and you strain to catch the words.	4.63	4.96	4.4	5
137	27	sound	ActualStimulus	246	heartbeat	4.83	4.65	5.07	5
138	27	text	ExplanatoryText	2120	You wash and rinse the dishes in the sink while your roommate vacuums.	5.36	4.06	5.68	5
139	53	sound	ActualStimulus	104	panting	4.96	5.37	5.06	5
140	53	text	ExplanatoryText	2222	You concentrate fully, hearing nothing, as you play your new video game.	6.34	5.31	6.28	5
141	54	word	ActualStimulus	957	razor	4.81	5.36	4.91	5
142	54	text	ExplanatoryText	7100	The telephone rings continuously as you look around the room to find it.	4.19	6.26	4.24	5
143	55	word	ActualStimulus	1004	swamp	5.14	4.86	5.29	5
144	55	text	ExplanatoryText	7250	You open the refrigerator and scan the shelves, searching for an evening snack.	5.74	4.63	5.9	5
145	51	word	ActualStimulus	434	theory	5.3	4.62	4.88	5

No	Pair	Type	Status	Code	Descr	Valence	Arousal	Dominance	Cluster
146	51	text	ExplanatoryText	2880	Your friend whispers to you in a meeting, and you strain to catch the words.	4.63	4.96	4.4	5
147	56	word	ActualStimulus	23	army	4.72	5.03	5.03	5
148	56	text	ExplanatoryText	8710	You are having a light workout today. After a warm-up, a thin layer of sweat begins to form. You stretch your quads and then start with some easy leg extensions.	6.39	4.79	6.72	5
149	57	word	ActualStimulus	966	rough	4.74	5.33	4.81	5
150	57	text	ExplanatoryText	9600	The nurse sinks the needle from the IV bag into your upper arm.	3.47	6.03	3.44	5

Appendix D

Comparing words, sounds, images and film clips

D.1 Study 3A Information Sheet

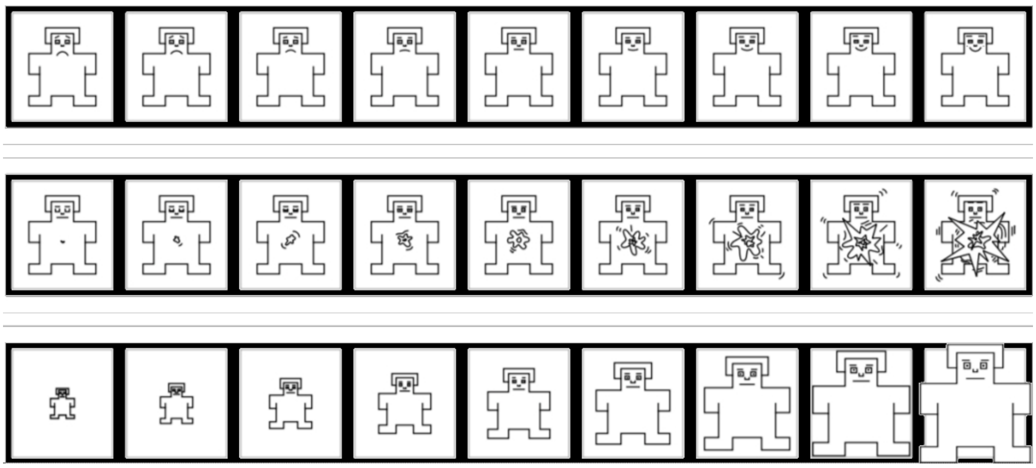
INFORMATION SHEET: Emotional ratings experiment

In this experiment, you will first respond to some questionnaires and then be played sequences of 4 types of stimuli: words, sounds, images and films. After each individual stimulus, you will be asked to rate how you would feel in real life if you experienced what it depicts. Please do not rate the stimulus in itself, but rather how you would feel if seeing/hearing it.

We will first run a practice trial for you to get more familiar with the experiment. Throughout the experiment you will receive instructions on-screen before the experimental tasks, but if anything is unclear, please do not hesitate to ask the researcher for details.

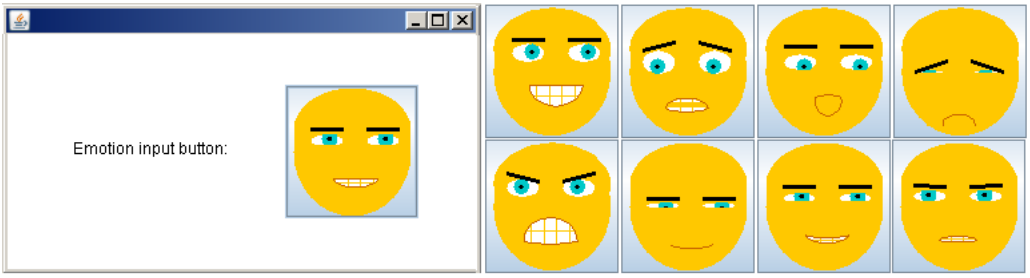
The rating procedure for each stimulus will involve two types of responses:

1. Below you have 3 sets of images, the first going from sad to happy, the second from bored/relaxed to alert and the third from submissive to dominant. From each set you will have to pick one image which best describes how you would feel in a situation related to the stimulus shown.



2. Below you have a cartoon figure which will change expression depending on how you move the mouse over it. When you think you have found an expression matching how you feel, make a click on that area of the face to record your response.

In addition, after each five images or films (but not words and sounds),



you will be asked to give the group an average rating in terms of how engaging and realistic the stimuli are etc.

D.2 List of 75 film stimuli

No	Description	YouTube Title	Cluster assigned	URL	T1 Time	T2 Time	T1 Frame	T2 Frame	Comments	Music
1	Board meeting	CTV Board of Directors Meeting 2/27/2014	1	https://www.youtube.com/watch?v=ezhnmjfQpqA	01:26:00	02:05:00	2580	3755	NA	0
2	Boy kick-boxing	9 year old kick boxing machine!	1	https://www.youtube.com/watch?v=qvI9MxND1zg	00:09:00	00:50:00	270	1510	NA	0
3	Catterpillar attacked by insect	Wheel bug (assassin bug) vs. silver-spotted skipper caterpillar	1	https://www.youtube.com/watch?v=N4v58Wg3u-A	00:00:00	00:40:00	0	1200	NA	1
4	Digging and sliding through a tunnel	Digging in Ogof Dydd Byraf	1	https://www.youtube.com/watch?v=z8uwpx_dIqE	00:06:00	00:46:00	200	1400	NA	0
5	Gum stuck on shoe	Gum stuck on shoe	1	https://www.youtube.com/watch?v=CsafMZbtuAU	00:10:00	00:50:00	300	1500	NA	0
6	Lecture on metal solidification	Metals and Alloys, lecture 5, Solidification	1	https://www.youtube.com/watch?v=9P0pZWHcmwA	00:28:00	01:09:00	700	1740	NA	0
7	Police make arrest	Newburgh police make arrest after high speed pursuit	1	https://www.youtube.com/watch?v=t0pptBn6308	00:02:00	00:42:00	60	1260	NA	0
8	Radio tuning	Tuning around the band march 27,2012 11:00 am EDT	1	https://www.youtube.com/watch?v=B-Z1dMvq28o	01:01:00	01:41:00	1840	3045	NA	0
9	Sheep blocking mountain road	Annoying some sheep in the Black Forrest on my BMW r1200gs	1	https://www.youtube.com/watch?v=HRRqYq90mN8	00:41:00	01:21:00	1230	2447	NA	0
10	Snake in a toilet	Rat snake in a toilet!	1	https://www.youtube.com/watch?v=ohq2ni7Xqmw	00:12:00	00:52:00	370	1565	NA	0
11	Spoiled child crying loudly	Zero Has a Tantrum	1	https://www.youtube.com/watch?v=TCzzv6fb5HA	00:50:00	01:30:00	1500	2700	NA	0
12	Spoiled child making demands	Moe is introduced to the concept of chores. Hes not a fan	1	https://www.youtube.com/watch?v=SXkbm6lkZoA	00:00:00	00:42:00	10	1280	NA	0
13	Student falling asleep in class	this is how boring college can be	1	https://www.youtube.com/watch?v=yYeyryadGcs	00:00:00	00:40:00	0	960	NA	0
14	Thunderstorm	Friday the 13th - September, 2013 - Louisiana Violent Thunderstorm Lightning Strikes Spooky	1	https://www.youtube.com/watch?v=3yFF3o6kMJI	00:00:00	00:40:00	20	1220	NA	0
15	Tombstones in cemetery	Bunhill Fields Burial Ground, London	1	https://www.youtube.com/watch?v=nfdNgHW18Z0	00:00:00	00:40:00	0	960	NA	0
16	Aggressive snake	Scary Macho Snake	2	https://www.youtube.com/watch?v=jT6qZpdGeIw	00:00:00	00:40:00	10	970	NA	0
17	Bee swarm attack	Man attacked by killer bees REAL VIDEO	2	https://www.youtube.com/watch?v=JEERv8sfLb0	00:21:00	01:01:00	530	1530	NA	0
18	Brain surgery	Corpus callosotomy.wmv	2	https://www.youtube.com/watch?v=GcyGqRk7zng	00:00:00	00:40:00	0	1200	NA	0
19	Car crash	Santa Monica Car Crash - Burning Car on 10 Freeway ??? 11/5/11	2	https://www.youtube.com/watch?v=Chqdywxadr4	01:30:00	02:10:00	2700	3920	NA	0

No	Description	YouTube Title	Cluster assigned	URL	T1 Time	T2 Time	T1 Frame	T2 Frame	Comments	Music
20	Dog attacks two girls on dust road	When Crazy Animals Attack Psychotic rabies dog attacks!!!!! Best Funny Animals 2014	2	https://www.youtube.com/watch?v=WHaTLVXRKiA	00:00:00	00:40:00	0	1225	A number of frames from around 00:34 - 00:36 (frames 1010 - 1100) were deleted due to very poor quality. The times/frames specified (T1/T2) apply after removing these poor-quality frames first.	0
21	Drainage of an abscess	got a bad back problem - abscess ID	2	https://www.youtube.com/watch?v=ZNdZvFLnOig	00:51:00	01:31:00	1530	2730	NA	0
22	Former school teacher evicted from his home	Baxter Jones Looses Home	2	https://www.youtube.com/watch?v=22iGeCl8JYY	00:02:00	00:42:00	50	1260	NA	0
23	Harbour storm and flood	Wick Harbour Storm	2	https://www.youtube.com/watch?v=abFQteCvQ_4	00:06:00	00:46:00	200	1400	NA	0
24	Man nearly run over by train	Crazy Sri Lankan Guy On the Railway Track - Narrow Escape	2	https://www.youtube.com/watch?v=OwRdvqjIhOo	00:16:00	00:56:00	385	1345	NA	0
25	Pipeline explosion	Otterburne pipeline explosion extended edition	2	https://www.youtube.com/watch?v=nlF97D6pQFU	01:10:00	01:50:00	2100	3300	NA	0
26	Street conflict with police	cops beating people up at occupy wall street	2	https://www.youtube.com/watch?v=xpOMlDVaXzc	00:00:00	00:40:00	0	1200	NA	0
27	Stressful traffic jam	CRAZY INDIAN TRAFFIC CONGESTION	2	https://www.youtube.com/watch?v=iEIk3RpV6RA	00:00:00	00:40:00	0	1020	NA	0
28	Syrian scene after massacre	[735] Syria: a Horrific Massacre against Civilians in Gas Station January 2, 2013 [Multilingual]	2	https://www.youtube.com/watch?v=Mkrg3xuLT4E	00:30:00	01:10:00	900	2120	NA	0
29	Violent man immobilised	Violent man with rabies is held down by 4 men!	2	https://www.youtube.com/watch?v=Xh82ZsLWd0s	00:30:00	01:10:00	750	1765	NA	0
30	War zone shooting	Syria, Iraq developments, December 11 2014	2	https://www.youtube.com/watch?v=vl3buFcHBuE	04:30:50	05:11:50	8115	9321	NA	0
31	Car rally	Camera Car Rally LANCIA DELTA S4 - 1?? ASSOLUTO	3	https://www.youtube.com/watch?v=pOTum1P2ljs	01:51:00	02:31:00	2780	3800	NA	1
32	Erotic scene with couple in bed	CENSORED. Contact researcher for details.	3	CENSORED. Contact researcher for details.	00:51:00	01:33:00	1300	2340	URL removed	1
33	Erotic scene with secretary	CENSORED. Contact researcher for details.	3	CENSORED. Contact researcher for details.	00:40:00	01:21:00	1000	2025	URL removed	1
34	Erotic scene: brunette with boyfriend	CENSORED. Contact researcher for details.	3	CENSORED. Contact researcher for details.	00:32:00	01:12:00	800	1818	URL removed	0
35	Erotic scene: couple by the pool	CENSORED. Contact researcher for details.	3	CENSORED. Contact researcher for details.	02:40:00	03:20:00	4001	5020	Top message and URL removed	1
36	Erotic scene: couple near pool table	CENSORED. Contact researcher for details.	3	CENSORED. Contact researcher for details.	01:36:00	02:18:00	2410	3452	URL removed	1

No	Description	YouTube Title	Cluster assigned	URL	T1 Time	T2 Time	T1 Frame	T2 Frame	Comments	Music
37	Erotic scene: man kissed across body by woman	CENSORED. Contact researcher for details.	3	CENSORED. Contact researcher for details.	00:35:00	01:15:00	875	1891	URL removed	1
38	Erotic scene: man kissed by woman in hotel	CENSORED. Contact researcher for details.	3	CENSORED. Contact researcher for details.	00:13:00	00:59:00	480	1480	Corporate logo was faded using Mplayer Delogo Filter	1
39	Fireworks	New Year Fireworks Display 2012 in Funchal, Madeira (HD)	3	https://www.youtube.com/watch?v=5XHvWJBgVJw	00:03:00	00:44:00	90	1090	NA	0
40	Football victory parade	St.Johnstone 2014 Scottish Cup Victory Parade. Perth High Street	3	https://www.youtube.com/watch?v=JDmQfnERRp0	00:32:00	01:12:00	960	2160	NA	0
41	Ibiza boat party	SHEIK N BEICK CLOSING PARTY IBIZA 2013	3	https://www.youtube.com/watch?v=qI_8pe6vqUE	00:25:00	01:06:00	605	1597	NA	0
42	Riding a roller-coaster	Gemini Rollercoaster Cedar Point Sandusky Ohio	3	https://www.youtube.com/watch?v=zP76UJ-FlaY	00:29:00	01:10:00	870	2100	NA	0
43	Skydiving	Saut parachute n??6 en PAC ??? Vannes Meucon	3	https://www.youtube.com/watch?v=IMA0ky2L4LQ	00:37:00	01:17:00	925	1940	NA	0
44	Surprise family reunion	Back Home !	3	https://www.youtube.com/watch?v=eAwfH8rJ_oc	01:35:00	02:16:00	2850	4095	NA	0
45	U2 Concert	U2 Concert - The Wave at Commonwealth Stadium	3	https://www.youtube.com/watch?v=nrlxgIzlSHI	00:09:00	00:49:00	270	1480	NA	0
46	Angkor Wat Temple in Cambodia	Inside Angkor Wat Temple in Siem Reap, Cambodia	4	https://www.youtube.com/watch?v=8pLMCHNHHgU	00:00:00	00:40:00	0	1220	NA	0
47	Artist playing the Earth Harp	William Close plays the Earth Harp at the Temple of Transition Burning Man 2011	4	https://www.youtube.com/watch?v=pIfeBL7vUcw	00:27:00	01:05:00	815	2016	Music happens to be playing in that environment, but is not aimed as background music for the video itself	1
48	Bird singing on a branch	Blackcap at RSPB Fowlmere - May 2014	4	https://www.youtube.com/watch?v=FtPWN2dy8sA	02:20:00	03:00:00	3515	4520	NA	0
49	Creating artistic hand lettering on post-its	Office Art: Hand Lettering	4	https://www.youtube.com/watch?v=wDOi3wK5-5I	00:09:00	00:50:00	240	1265	NA	1
50	Flowers and tulips	Flower and Tulip DVD - Flowers of Holland - Videos For Relaxation	4	https://www.youtube.com/watch?v=b9NqbU5JpqM	00:25:00	01:05:00	740	1950	NA	1
51	Heart-warming family event	Amazing little video... Habitat ribbon cutting for Shayna Conway's family	4	https://www.youtube.com/watch?v=_58lfv8yNIQ	00:33:00	01:13:00	991	2220	NA	1
52	Japanese-style garden	The Pagoda Garden - Norfolk, VA	4	https://www.youtube.com/watch?v=BKZdy9Ga0S4	00:02:00	00:40:00	25	1030	NA	1

No	Description	YouTube Title	Cluster assigned	URL	T1 Time	T2 Time	T1 Frame	T2 Frame	Comments	Music
53	Piano on display in city square	123 Piano in City Pavilion Ghent Belgium Part 1	4	https://www.youtube.com/watch?v=QrB8GXoxo8A	00:15:00	00:56:00	450	1700	Music happens to be playing in that environment, but is not aimed as background music for the video itself	1
54	Puppy barks softly	Cute Puppy talking in his Sleep	4	https://www.youtube.com/watch?v=QsSvl2FNcKo	00:00:00	00:40:00	0	1199	NA	0
55	Relaxing yoga workout	Fitness Yoga Workout Episode 2	4	https://www.youtube.com/watch?v=2HAFhDuKVvQ	00:05:00	00:46:00	170	1380	NA	1
56	Seagulls flying over beach	Seagulls... Sea Sun Sand Surf ! Birds at Newgale Beach Pembrokeshire Wales UK British Britain	4	https://www.youtube.com/watch?v=zsBQE9a7HgY	00:00:00	00:41:00	0	1040	NA	0
57	Strawberry cheesecake recipe	Strawberry Cheesecake Dessert Shooter Recipe	4	https://www.youtube.com/watch?v=gn1j7xOBewg	01:33:00	02:15:00	2800	4050	NA	1
58	Tropical island scenes	Vomo Island Fiji, Vomo. It's Personal	4	https://www.youtube.com/watch?v=YVcEXF7aL98	00:02:00	00:42:00	30	1016	NA	1
59	Venice in the evening	One Evening in Campo Santa Maria Nova, Venice	4	https://www.youtube.com/watch?v=d-0tv_4mtJY	00:04:00	00:45:00	120	1340	NA	1
60	Waterfalls in forest	Waterfalls – A Short Visit to the Catlins	4	https://www.youtube.com/watch?v=hyb5lpBHUuE	00:05:00	00:46:00	120	1157	NA	1
61	Battleship open to visitors	Battleship Missouri Memorial	5	https://www.youtube.com/watch?v=SnaWFg7XF84	02:50:00	03:30:00	5100	6315	NA	0
62	Dog panting on floor	Border Collie panting	5	https://www.youtube.com/watch?v=3xALJkLtb04	00:35:00	01:15:00	840	1800	NA	0
63	Grizzly bear and then wolf crossing forest	wolves traveling with grizzly bear	5	https://www.youtube.com/watch?v=159Q_RTsX_o	00:07:00	00:47:00	230	1450	NA	0
64	Knife displayed	Enlan EL-08 (2/2)	5	https://www.youtube.com/watch?v=Xr-FDYakJQo	01:20:00	02:00:00	2400	3600	NA	0
65	Lecture on the temporal discounting of reward	Lecture 18 (Lagrange Multipliers)	5	https://www.youtube.com/watch?v=HKGQ-dKzdMU	03:25:00	04:05:00	6163	7353	NA	0
66	Medical assessment	Cardiothoracic – Physical Assessment	5	https://www.youtube.com/watch?v=qqyTY8_vtB4	04:44:00	05:25:00	7100	8130	NA	0
67	Military training camp with obstacle run	MilitaryTrainingCamps.com - The NastyNick Course at the United States Army John F. Kennedy Special	5	https://www.youtube.com/watch?v=N3Q46iaJrX8	00:13:00	00:54:00	395	1623	NA	0
68	Pig farm	Pig farm Industrail Piggery - Free HD Stock Footage	5	https://www.youtube.com/watch?v=SA_qmxYJWUA	00:00:00	00:40:00	0	1000	NA	0
69	River with cicadas buzzing	Anclote river - Starkey Wilderness - Cicadas buzzing	5	https://www.youtube.com/watch?v=q-f376XyvgU	00:49:00	01:29:00	1470	2670	NA	0

No	Description	YouTube Title	Cluster assigned	URL	T1 Time	T2 Time	T1 Frame	T2 Frame	Comments	Music
70	Scuba diving into a lake with jellyfish	Kakaban Jellyfish Lake	5	https://www.youtube.com/watch?v=mqfXniAm7ck	00:27:00	01:07:00	775	2006	NA	1
71	Speedboat crossing choppy waters	Speedboat in rough seas	5	https://www.youtube.com/watch?v=ZKFF5Ud_7OO	00:25:00	01:05:00	750	1960	NA	0
72	Street in Athens	Athens walk ??? 3rd route - Agias Filotheis street	5	https://www.youtube.com/watch?v=VoLbLIpF6-c	00:04:00	00:46:00	160	1380	NA	0
73	Tense dialogue	Rough time	5	https://www.youtube.com/watch?v=Xz9nhcpuO3o	01:00:00	01:40:00	1510	2525	NA	0
74	Train passing by	SD70MAC led coal train climbs Mullan Pass	5	https://www.youtube.com/watch?v=In3zXS4dKQA	00:45:00	01:25:00	1350	2563	NA	0
75	University basketball game	Prelaunch - The North Greenville University Exhibition Game-The Relaunching Of Paladin Basketball	5	https://www.youtube.com/watch?v=K_aucKnQXcw	05:11:00	05:52:00	9344	10570	Music happens to be playing in that environment, but is not aimed as background music for the video itself	1

D.3 Stimulus quartets: matched images, words and sounds, with films

A selection of 37 films were matched to words, sounds, and images within the various clusters, as shown below. A subset of 25 of these were selected for further analysis, given that 12 of the films were excluded due to implausibility and 'unusual' PAD values. Any films excluded have been marked below using an ✗ sign, and problematic PAD values are highlighted in grey:

No	Cluster	Type	Database code	Description	Mean Valence	Mean Arousal	Mean Dominance
1	Neutral	word	1004	swamp	3.83	3.38	3.65
2	Neutral	word	23	army	3.05	4.62	3.60
3	Neutral	word	434	theory	4.85	3.70	4.38
4	Neutral	word	957	razor	3.52	4.00	4.03
5	Neutral	word	966	rough	3.33	3.83	3.83
6	Neutral	film	NA	Battleship open to visitors	4.43	3.03	3.90
7	Neutral	film	NA	Dog panting on floor	4.98	3.05	4.43
8	Neutral	film	NA	Knife displayed	3.73	4.23	3.97
9	Neutral	film	NA	Lecture on the temporal discounting of reward	4.62	3.00	3.83
10	Neutral	film	NA	Medical assessment	3.60	3.12	3.55
11	Neutral	film	NA	✗ Pig farm [a]	2.47	3.85	3.58
12	Neutral	film	NA	✗ River with cicadas buzzing [b]	4.30	2.75	3.72
13	Neutral	film	NA	✗ Street in Athens [b]	4.97	2.98	4.58
14	Neutral	film	NA	✗ Tense dialogue [a]	2.75	4.52	3.77
15	Neutral	film	NA	✗ Train passing by [b]	4.42	2.92	3.80
16	Neutral	image	1645	Wolf	4.37	5.02	3.53
17	Neutral	image	1908	Jellyfish	4.10	4.37	3.47
18	Neutral	image	2220	MaleFace	3.33	4.33	3.70
19	Neutral	image	3550.2	Coach	3.50	4.13	3.53
20	Neutral	image	9422	Battleship	3.78	3.98	3.48
21	Neutral	sound	104	Panting	4.02	4.67	3.98
22	Neutral	sound	130	Pig	4.12	3.72	4.18
23	Neutral	sound	246	HeartBeat	4.35	4.48	3.95
24	Neutral	sound	425	Train	4.47	3.92	3.83
25	Neutral	sound	722	Walking	4.05	4.07	3.98
26	Mildly negative	word	405	solemn	3.65	3.03	3.62
27	Mildly negative	word	608	skull	3.33	4.05	3.53
28	Mildly negative	word	806	immature	3.03	4.00	3.78
29	Mildly negative	word	834	kick	3.45	4.28	3.93
30	Mildly negative	word	913	obnoxious	2.82	4.33	3.82
31	Mildly negative	film	NA	Catterpillar attacked by insect	1.98	5.63	3.07

32	Mildly negative	film	NA	Gum stuck on shoe	4.42	3.65	4.30
33	Mildly negative	film	NA	Radio tuning	3.70	2.72	3.90
34	Mildly negative	film	NA	Snake in a toilet	1.78	6.17	2.50
35	Mildly negative	film	NA	Spoiled child making demands	4.00	3.88	4.33
36	Mildly negative	image	1230	Spider	2.68	5.10	3.10
37	Mildly negative	image	1240	Spider	2.65	5.38	2.98
38	Mildly negative	image	2206	Fingerprint	3.52	3.90	3.25
39	Mildly negative	image	2410	Boy	3.17	4.35	3.72
40	Mildly negative	image	9210	Rain	3.45	3.65	3.50
41	Mildly negative	sound	358	Writing	3.98	4.02	3.70
42	Mildly negative	sound	382	Shovel	3.52	3.70	3.62
43	Mildly negative	sound	700	Toilet	4.08	2.58	3.88
44	Mildly negative	sound	708	Clock	3.78	2.92	3.63
45	Mildly negative	sound	723	Radio	3.72	3.65	4.17
46	Very negative	word	15	ambulance	2.38	4.85	2.90
47	Very negative	word	591	drown	1.28	5.40	2.17
48	Very negative	word	604	scared	2.23	5.20	2.35
49	Very negative	word	616	trauma	1.67	4.97	2.40
50	Very negative	word	618	victim	1.85	4.78	2.47
51	Very negative	film	NA	Car crash	1.67	5.13	2.75
52	Very negative	film	NA	Drainage of an abscess	1.22	6.00	2.20
53	Very negative	film	NA	Street conflict with police	1.57	6.22	3.12
54	Very negative	film	NA	✗ Stressful traffic jam [c]	3.08	4.83	3.43
55	Very negative	film	NA	Syrian scene after massacre	1.00	6.52	2.05
56	Very negative	film	NA	Violent man immobilised	1.18	6.03	2.30
57	Very negative	image	3110	BurnVictim	0.72	5.75	2.00
58	Very negative	image	3130	Mutilation	0.65	5.92	2.02
59	Very negative	image	9183	HurtDog	0.78	5.48	2.60
60	Very negative	image	9187	InjuredDog	1.08	5.38	2.83
61	Very negative	image	9412	DeadMan	0.78	5.47	2.33
62	Very negative	sound	115	Bees	2.47	5.53	2.42
63	Very negative	sound	279	Attack1	0.60	6.47	2.53
64	Very negative	sound	292	MaleScream	1.38	5.92	2.35
65	Very negative	sound	420	CarHorns	2.60	4.82	3.35
66	Very negative	sound	424	CarWreck	1.45	6.20	2.32
67	Positive exciting	word	152	excitement	6.47	5.82	5.47
68	Positive exciting	word	384	sex	6.00	5.50	5.00
69	Positive exciting	word	422	surprised	5.57	5.38	4.18
70	Positive exciting	word	512	erotic	5.45	5.15	4.52

71	Positive exciting	word	530	sexy	5.87	4.82	5.03
72	Positive exciting	film	NA	Erotic scene: brunette with boyfriend	5.75	5.15	4.77
73	Positive exciting	film	NA	Erotic scene: couple by the pool	5.57	5.18	4.80
74	Positive exciting	film	NA	Erotic scene: couple near pool table	5.63	5.07	4.77
75	Positive exciting	film	NA	✗ Erotic scene: man kissed across body by woman [b]	5.62	4.82	4.82
76	Positive exciting	film	NA	Erotic scene: man kissed by woman in hotel	6.00	5.30	4.97
77	Positive exciting	film	NA	Erotic scene with couple in bed	5.77	5.03	4.97
78	Positive exciting	film	NA	✗ Erotic scene with secretary [c]	4.98	4.45	4.32
79	Positive exciting	film	NA	✗ Football victory parade [b]	5.80	4.58	4.60
80	Positive exciting	image	4659	EroticCouple	5.43	5.32	4.57
81	Positive exciting	image	4668	EroticCouple	5.60	5.27	4.62
82	Positive exciting	image	4670	EroticCouple	5.52	5.03	4.50
83	Positive exciting	image	8200	WaterSkier	5.63	5.35	4.60
84	Positive exciting	image	8492	Rollercoaster	5.57	6.22	3.50
85	Positive exciting	sound	201	EroticFem1	5.42	5.25	4.85
86	Positive exciting	sound	311	Crowd2	5.63	5.02	4.72
87	Positive exciting	sound	360	RollerCoaster	4.12	5.27	3.65
88	Positive exciting	sound	815	RockNRoll	6.18	4.70	5.28
89	Positive exciting	sound	817	Bongos	6.10	4.35	4.92
90	Positive serene	word	320	politeness	5.57	2.92	4.48
91	Positive serene	word	466	useful	5.88	3.55	5.03
92	Positive serene	word	761	garden	5.90	2.53	4.62
93	Positive serene	word	786	heal	6.00	3.67	4.52
94	Positive serene	word	796	humane	5.48	2.98	4.62
95	Positive serene	film	NA	Bird singing on a branch	5.47	2.32	4.37
96	Positive serene	film	NA	✗ Creating artistic hand lettering on post-its [a]	4.88	2.88	4.32
97	Positive serene	film	NA	Heart-warming family event	5.68	3.23	4.32

98	Positive serene	film	NA	Piano on display in city square	5.70	3.18	4.42
99	Positive serene	film	NA	✗ Puppy barks softly [d]	6.68	3.27	4.88
100	Positive serene	film	NA	Relaxing yoga workout	5.30	2.45	4.37
101	Positive serene	film	NA	Seagulls flying over beach	5.53	2.87	4.45
102	Positive serene	film	NA	✗ Strawberry cheesecake recipe [e]	5.73	3.43	4.97
103	Positive serene	image	2156	Family	5.85	3.17	4.57
104	Positive serene	image	2314	Binoculars	6.20	3.30	4.68
105	Positive serene	image	5200	Flowers	5.37	2.45	4.45
106	Positive serene	image	7390	IceCream	5.50	3.10	4.65
107	Positive serene	image	7480	Pasta	6.37	3.67	5.05
108	Positive serene	sound	112	Kids1	5.42	3.52	4.60
109	Positive serene	sound	150	Seagull	5.15	3.25	4.28
110	Positive serene	sound	151	Robin	5.83	2.67	4.50
111	Positive serene	sound	809	Harp	5.75	2.93	4.55
112	Positive serene	sound	810	Beethoven	6.12	3.22	4.63

[a] Average Valence rating is below the level expected within the cluster.
[b] Average Arousal rating is below the level expected within the cluster.
[c] Discrepancy on all three PAD dimensions, relative to the other values in the cluster.
[d] Average Valence rating is above the level expected within the cluster.
[e] Among this cluster's film clips, this presented the highest Arousal - which was judged inconsistent with a 'serene' cluster.

D.4 R session information

```
R version 3.4.0 (2017-04-21)
Platform: x86_64-pc-linux-gnu (64-bit)
Running under: Ubuntu 16.04.2 LTS

Matrix products: default
BLAS: /usr/lib/libblas/libblas.so.3.6.0
LAPACK: /usr/lib/lapack/liblapack.so.3.6.0

locale:
 [1] LC_CTYPE=en_GB.UTF-8       LC_NUMERIC=C               LC_TIME=en_GB.UTF-8        LC_COLLATE=en_GB.UTF-8     LC_MONETARY=en_GB.UTF-8
 [6] LC_MESSAGES=en_GB.UTF-8    LC_PAPER=en_GB.UTF-8       LC_NAME=C                  LC_ADDRESS=C               LC_TELEPHONE=C
[11] LC_MEASUREMENT=en_GB.UTF-8 LC_IDENTIFICATION=C

attached base packages:
[1] grid      stats     graphics  grDevices utils     datasets  methods   base

other attached packages:
 [1] texreg_1.36.23      stargazer_5.2       optmatch_0.9-7      survival_2.41-3     MCMCglmm_2.24       ape_4.1
 [7] coda_0.19-1         lme4_1.1-13         Matrix_1.2-10       paran_1.5.1         MASS_7.3-47         latticeExtra_0.6-28
[13] RColorBrewer_1.1-2  gridExtra_2.2.1     ggplot2_2.2.1       psych_1.7.5         tidyr_0.6.3         plyr_1.8.4
[19] data.table_1.10.4   stringr_1.2.0       xtable_1.8-2        lattice_0.20-35

loaded via a namespace (and not attached):
 [1] Rcpp_0.12.11     compiler_3.4.0   nloptr_1.0.4     RItools_0.1-15   tools_3.4.0      digest_0.6.12    tibble_1.3.3
 [8] gtable_0.2.0     nlme_3.1-131     rlang_0.1.1      parallel_3.4.0   SparseM_1.77     foreign_0.8-67   tensorA_0.36
[15] minqa_1.2.4      corpcor_1.6.9    magrittr_1.5     scales_0.4.1     splines_3.4.0    svd_0.4          abind_1.4-5
[22] mnormt_1.5-5     colorspace_1.3-2 cubature_1.3-8   stringi_1.1.5    lazyeval_0.2.0   munsell_0.4.3
```

Appendix E

Comparing film clips and (immersive or non-immersive) VEs

E.1 Listing films and VEs used

No	Block	Cluster	ResearcherDescription	Second LifeURLTag	Parcel	Region	Rating	URL	Type	Music	Session
1	1	1	Spoiled child making demands		NA	NA	NA	See previous Appendix	film	0	1
2	1	1	Snake in a toilet		NA	NA	NA	See previous Appendix	film	0	1
3	1	1	Caterpillar attacked by insect		NA	NA	NA	See previous Appendix	film	1	1
4	1	1	Radio tuning		NA	NA	NA	See previous Appendix	film	0	1
5	1	1	Gum stuck on shoe		NA	NA	NA	See previous Appendix	film	0	1
6	2	1	Dark junkyard against modern lit-up buildings in the background		New Tokyo	FireStorm	Moderate	http://maps.secondlife.com/secondlife/FireStorm/141/218/29	VE	0	1
7	2	1	Narrow street between pubs, paved with cubic stone, in port town		Noctis	Port Babbage	Moderate	http://maps.secondlife.com/secondlife/Port%20Babbage/63/206/106	VE	1	1
8	2	1	Desolate concrete buildings on abandoned industrial estate, with some tumbleweed and old newspapers flying in the wind		Alien Beta	Harshap	Moderate	http://maps.secondlife.com/secondlife/Harshap/149/111/102	VE	0	1
9	2	1	Road with rusty old vehicles parked alongside it, similar to a scrap yard		Random Labs	Sky City	Moderate	http://maps.secondlife.com/secondlife/Sky%20City/198/196/35	VE	0	1
10	2	1	Dirty back yard filled with trash and junk		Home Sweet Home	Tariah	Adult	http://maps.secondlife.com/secondlife/Tariah/220/26/2502	VE	0	1
11	3	1	Rainy alley between buildings, with trash and old boxes lying around		Rainy Alley	Bay City - Sconset	General	http://maps.secondlife.com/secondlife/Bay%20City%20-%20Sconset/171/207/1614	VE	0	2
12	3	1	Uninviting and creepy circus tents		Carnival of Chaos	The Wash	General	http://maps.secondlife.com/secondlife/The%20Wash/187/221/23	VE	1	2
13	3	1	Swamp with abandoned stilt houses and shacks, and trash littered around		Muddy Garbage Swamp	Agriopis	Moderate	http://maps.secondlife.com/secondlife/Agriopis/180/118/23	VE	0	2
14	3	1	Exhibition hall / shop with picture frames, furniture and trinkets		Gift Shop	Furniture	Moderate	http://maps.secondlife.com/secondlife/Furniture/48/113/39	VE	0	2
15	3	1	Steampunk, inhospitable city environment		Grunge City	Isle of Tharen	Moderate	http://maps.secondlife.com/secondlife/Isle%20of%20Tharen/101/195/22	VE	0	2

No	Block	Cluster	ResearcherDescription	Second LifeURLTag	Parcel	Region	Rating	URL	Type	Music	Session
16	4	1	Large grey cathedral, with grand entrance and a fountain in front of it. Heavy metal doors open to reveal two rows of pews, stained glass at the far side, all with an overall solemn air. A beggar lies at the entrance.		Basilica St. Peter	Mediterraneo OC	Adult	http://maps.secondlife.com/secondlife/Mediterraneo%20OC/47/27/25	VE	0	2
17	4	1	Desert area, with cliffs, dunes, and very small weed bushes. Scraps of metal and a trailer are also present, as well as some abandoned fun rides.		Tableau	Tableau	General	http://maps.secondlife.com/secondlife/Tableau/119/185/23	VE	0	2
18	4	1	Fairly dull-looking museum, with various clock models on display, ticking away.		The Clock Museum	Triglav	Moderate	http://maps.secondlife.com/secondlife/Triglav/94/35/75	VE	0	2
19	4	1	Grassy field with a large number of snakes and spiders moving around.		Pet Snake and Pet Spider	Jupiter	Moderate	http://maps.secondlife.com/secondlife/Jupiter/28/84/27	VE	1	2
20	4	1	Reddish skies over abandoned circus / carnival. Overall looks dark and foreboding.		The Sandcastle	Furor	Moderate	http://maps.secondlife.com/secondlife/Furor/123/60/2994	VE	0	2
21	5	2	Car crash		NA	NA	NA	See previous Appendix	film	0	1
22	5	2	Street conflict with police		NA	NA	NA	See previous Appendix	film	0	1
23	5	2	Drainage of an abscess		NA	NA	NA	See previous Appendix	film	0	1
24	5	2	Syrian scene after massacre		NA	NA	NA	See previous Appendix	film	0	1
25	5	2	Violent man immobilised		NA	NA	NA	See previous Appendix	film	0	1
26	6	2	Morgue with examination tables and mortuary cabinets		AS Forensic and Pathology Lab	Baraka Point	Adult	http://maps.secondlife.com/secondlife/Baraka%20Point/32/156/22	VE	0	1
27	6	2	Horror world where monsters run after players, and items levitate in haunted spaces		Village of the Damned	Twilight Hollow	General	http://maps.secondlife.com/secondlife/Twilight%20Hollow/18/240/3930	VE	0	1
28	6	2	Dark cave / volcano with corpses hung up for torture, with some explicit content as well		CENSORED. Contact researcher for details.	Hell	Moderate	http://maps.secondlife.com/secondlife/Hell/65/92/32	VE	1	1
29	6	2	Halloween-themed dark space, with haunted mansion		Halloween Town	Duel One	General	http://maps.secondlife.com/secondlife/Duel%20One/171/91/24	VE	1	1
30	6	2	Dirty rooms and torture chambers in abandoned morgue, near scary fun fair, with freak show		Boardwalk City	Weedon Island	Moderate	http://maps.secondlife.com/secondlife/Weedon%20Island/224/155/36	VE	0	1

No	Block	Cluster	ResearcherDescription	Second LifeURLTag	Parcel	Region	Rating	URL	Type	Music	Session
31	7	2	Dark catacombs in a jungle, with many narrow corridors and tunnels – similar in style to South American ancient ruins		Pino	Pino	Moderate	http://maps.secondlife.com/secondlife/Pino/210/72/912	VE	0	2
32	7	2	Dark streets in small ransacked town. Dead bodies and garbage lying on the ground. After a short while zombies emerge and chase the player		Hunt for the Undead	Nuvoletta	Moderate	http://maps.secondlife.com/secondlife/Nuvoletta/9/250/2001	VE	0	2
33	7	2	Zombie-infested, gory subway and neighbouring buildings.		Psycho City	Turia	Adult	http://maps.secondlife.com/secondlife/Turia/11/230/24	VE	1	2
34	7	2	Haunted mansion with chilling background creaks and groans, monsters, and anatomical curiosities on display		The Hum and the Shiver	Picklemoon	Moderate	http://maps.secondlife.com/secondlife/Picklemoon/225/199/25	VE	0	2
35	7	2	Very dark, misty forest / garden in front of mansion, with graves, complete with voices, the headless horseman, ghosts and skeletons. The entrance to the mansion is littered with bloody razors.		Haunted Heightened Passion	Caymen Shores	Adult	http://maps.secondlife.com/secondlife/Caymen%20Shores/195/152/21	VE	1	2
36	8	2	Gory scene in slaughterhouse, with morbidly obese butcher. Outside there is a small playground, with blood splatter on swings, slides and a roundabout.		Immortal Hearts Family	Burning Hart	Moderate	http://maps.secondlife.com/secondlife/Burning%20Hart/18/84/24	VE	1	2
37	8	2	Eerie autumn forest, with haunted mansion. It is the scene of ghosts floating midair, hooded figures performing rituals, and dead bodies		Pickled Spooky	PickeSong	Moderate	http://maps.secondlife.com/secondlife/PickleSong/27/23/41	VE	0	2
38	8	2	Very dark and grim environment, with blackened brick walls enclosing a staircase, which leads to a graveyard and a haunted mansion. Crucified corpses and skeletons are present, as well as skinned animals and giant spiders.		The Forsaken	Full Throttle	Moderate	http://maps.secondlife.com/secondlife/Full%20Throttle/67/180/29	VE	1	2

No	Block	Cluster	ResearcherDescription	Second LifeURLTag	Parcel	Region	Rating	URL	Type	Music	Session
39	8	2	Large foreboding swamp, bordering on some wooden stilt houses. Skulls on a spike mark the entrance to one of these houses.		Dolls of Death	Mlastina de Anticii	Adult	http://maps.secondlife.com/secondlife/Mlastina%20de%20Anticii/152/161/30	VE	0	2
40	8	2	Barren landscape, with eerie circus caravans		Desolation Island	Bahiana	Adult	http://maps.secondlife.com/secondlife/Bahiana/115/147/1503	VE	0	2
41	9	3	Erotic scene with couple in bed		NA	NA	NA	See previous Appendix	film	1	1
42	9	3	Erotic scene: brunette with boyfriend		NA	NA	NA	See previous Appendix	film	0	1
43	9	3	Erotic scene: couple near pool table		NA	NA	NA	See previous Appendix	film	1	1
44	9	3	Erotic scene: man kissed by woman in hotel		NA	NA	NA	See previous Appendix	film	1	1
45	9	3	Erotic scene: couple by the pool		NA	NA	NA	See previous Appendix	film	1	1
46	10	3	Strip club on exotic beach, with some nudity		Hot Dreams Strip Club	Sexy Sands	Adult	http://maps.secondlife.com/secondlife/Sexy%20Sands/159/153/24	VE	0	2
47	10	3	Red-light district street, full of strip clubs and specialist shops		Redlight Escorts	CENSORED. Contact researcher for details.	Adult	CENSORED. Contact researcher for details.	VE	0	2
48	10	3	Main square in town, from which various streets lead to strip clubs		Escort Oasis	Escort Oasis	Adult	http://maps.secondlife.com/secondlife/escort%20oasis/105/71/31	VE	0	2
49	10	3	Firework shop with demonstrations of colourful and fun fireworks		BG's Fireworks	Xalfor	General	http://maps.secondlife.com/secondlife/Xalfor/81/129/21	VE	0	2
50	10	3	Disneyland-like environment, with various rides, tours, ponies, trams and pink castle in the background.		Magicland Castle	Bracket	Moderate	http://maps.secondlife.com/secondlife/Bracket/180/133/30	VE	1	2
51	11	3	Amusement park with carousel and Ferris wheel, surrounded by some houses and the sea		Sea Breeze Art and Carnival	Phantomn District [sic!]	Moderate	http://maps.secondlife.com/secondlife/Phantomn%20District/234/75/22	VE	0	1
52	11	3	Street in the red-light district, with strip bars and explicit content		Le Folie's Pigalle	Madeley	Adult	http://maps.secondlife.com/secondlife/Madeley/229/61/2016	VE	0	1
53	11	3	Open-air strip club with some nudity		The Place	Guerreiros	Adult	http://maps.secondlife.com/secondlife/GUERREIROS/168/226/27	VE	0	1

No	Block	Cluster	ResearcherDescription	Second LifeURLTag	Parcel	Region	Rating	URL	Type	Music	Session
54	11	3	Colourful bar with arcade games lined up against every wall		Back To The 80s Arcade	Kindred Spirit	Moderate	http://maps.secondlife.com/secondlife/Kindred%20Spirit%201/75/91/22	VE	0	1
55	11	3	Sea-side roller coaster and ticketing booth		MuSe IsLe Rollercoaster at Metanomics Island	Metanomics Island	Moderate	http://maps.secondlife.com/secondlife/Metanomics/171/43/32	VE	1	1
56	12	3	Strip club on exotic beach, surrounded by the sea, rock cliffs and trees		LuvLace Strip Club	Sexy Sands	Adult	http://maps.secondlife.com/secondlife/Sexy%20Sands/96/76/25	VE	0	1
57	12	3	Bar with arcade games, bowling lanes, pool tables, and snack / popcorn bar		Galaxy Bowling and Arcade Games	Dreyfus	General	http://maps.secondlife.com/secondlife/Dreyfus/86/72/1502	VE	1	1
58	12	3	Beach with palm trees, bars and dance / strip clubs		Castle Nights Dance Club	Buttercup Island	Adult	http://maps.secondlife.com/secondlife/Buttercup%20Island/45/178/22	VE	0	1
59	12	3	Explicit content: (semi-)nude avatars, adult scenery and artwork (sculptures) in / outside a bar, against cityscape		CENSORED. Contact researcher for details.	Secret Sands	Adult	CENSORED. Contact researcher for details.	VE	0	1
60	12	3	Open-air strip club with some nudity		CENSORED. Contact researcher for details.	Brothel	Adult	CENSORED. Contact researcher for details.	VE	0	1
61	13	4	Piano on display in city square		NA	NA	NA	See previous Appendix	film	1	2
62	13	4	Bird singing on a branch		NA	NA	NA	See previous Appendix	film	0	2
63	13	4	Heart-warming family event		NA	NA	NA	See previous Appendix	film	1	2
64	13	4	Seagulls flying over beach		NA	NA	NA	See previous Appendix	film	0	2
65	13	4	Relaxing yoga workout		NA	NA	NA	See previous Appendix	film	1	2
66	14	4	Rich, Arabian-inspired interior, with arabesque motifs decorating the walls, mosaic floors and small pool		Ally Adventure	Coral Sands	Adult	http://maps.secondlife.com/secondlife/Coral%20Sands/225/13/23	VE	0	1
67	14	4	Large garden in front of sumptuous palace, with luxurious entrance and interiors		Wilanow Royal Palace and Gardens	Oceanea	Moderate	http://maps.secondlife.com/secondlife/Oceanea/155/148/30	VE	0	1
68	14	4	Large dining area, with a variety of dishes on display, covering several long tables		Food Connection	Depoz Specialities	General	http://maps.secondlife.com/secondlife/Depoz%20Specialties/254/158/27	VE	0	1

No	Block	Cluster	ResearcherDescription	Second LifeURLTag	Parcel	Region	Rating	URL	Type	Music	Session
69	14	4	Rustic, French-inspired town, with old buildings, many gardens, terraces and arches		Ville de Coeur	Aquitaine Coeur Nord	Moderate	http://maps.secondlife.com/secondlife/Aquitaine%20Coeur%20Nord/181/93/34	VE	0	1
70	14	4	Sea-side resort, with cosy cabins and wooden huts facing the sea		Isle of Breezes	Shipton	Moderate	http://maps.secondlife.com/secondlife/Shipton/213/195/22	VE	0	1
71	15	4	Toy store, full of colourful items – colouring books, climbing tunnels for children, coin-operated toy-grabbing machines etc.		Spronkwings	Misthaven	Moderate	http://maps.secondlife.com/secondlife/Misthaven/222/46/22	VE	0	2
72	15	4	Immense royal palace, complete with baroque decorations and a large entrance guarded by two stone lions on each side. The interior is particularly opulent, with lush carpets, sculptures, and rosewood-covered walls.		The Rose Theatre and Art Gallery	Angel Manor	Moderate	http://maps.secondlife.com/secondlife/Angel%20Manor/3/185/21	VE	0	2
73	15	4	Replica of Venice afloat on water, with gondolas, romantic-looking buildings, and restaurants facing the water.		Venice	Yumix Prada	Moderate	http://maps.secondlife.com/secondlife/Yumix%20Prada/115/87/30	VE	0	2
74	15	4	Rocky seafront, with lush vegetation, flowers and fountains, as well as wooden trellises for supporting plants. A sandy beach lies below the seafront.		Garden Plants	Mountains of Creta	General	http://maps.secondlife.com/secondlife/Mountains%20of%20Creta/126/101/33	VE	0	2
75	15	4	Rocky beach, with sea waves washing up along the shore. There is a lighthouse nearly, as well as a small campfire and some sun beds with cushions. Some stairs in the background lead up to a small villa.		Black Basalt Beach	Brandy Wine Island	Moderate	http://maps.secondlife.com/secondlife/Brandy%20Wine%20Island/93/92/21	VE	0	2
76	16	4	Mediterranean landscape with bright sun, lush vegetation and flowers, and a stone bridge over a body of water, extending into the sea in the distance. A Greek-style, open-air amphitheatre is also present.		Da Vinci Gardens	Kalepa	General	http://maps.secondlife.com/secondlife/Kalepa/182/192/25	VE	0	2
77	16	4	Relaxing oriental / Indian gardens with stone-paved paths and small stone temples.		Yoga Mediation	Quietly Tuesday	Moderate	http://maps.secondlife.com/secondlife/Quietly%20Tuesday/164/132/33	VE	0	2

No	Block	Cluster	ResearcherDescription	Second LifeURLTag	Parcel	Region	Rating	URL	Type	Music	Session
78	16	4	Lush garden filled with colourful flowers, bushes and trees.		Martie's Elegant Plants	Cavettaz	Moderate	http://maps.secondlife.com/secondlife/Cavettaz/198/70/35	VE	0	2
79	16	4	Large gothic castle / fortress, in forest. Nearby there is a river and a fountain, as well as stone decorations.		Medieval manor house	Kismet Northwinds	Moderate	http://maps.secondlife.com/secondlife/Kismet%20Northwinds/103/134/23	VE	0	2
80	16	4	Luxurious patisserie and gelato place near the sea.		Patisserie De Bauchery	Left Hand Column	Moderate	http://maps.secondlife.com/secondlife/Left%20Hand%20Column/125/84/29	VE	0	2
81	17	5	Lecture on the temporal discounting of reward		NA	NA	NA	See previous Appendix	film	0	2
82	17	5	Battleship open to visitors		NA	NA	NA	See previous Appendix	film	0	2
83	17	5	Dog panting on floor		NA	NA	NA	See previous Appendix	film	0	2
84	17	5	Knife displayed		NA	NA	NA	See previous Appendix	film	0	2
85	17	5	Medical assessment		NA	NA	NA	See previous Appendix	film	0	2
86	18	5	Steampunk, grey world, with old and abandoned factories and buildings.		Prefabrica	Prefabrica	General	http://maps.secondlife.com/secondlife/PREFABRICA/134/217/46	VE	0	2
87	18	5	Pier next to several stationed battleships		Battleship Hub	Tulagi	Moderate	http://maps.secondlife.com/secondlife/Tulagi/1/220/24	VE	0	2
88	18	5	Beach with many cute dogs roaming around, with small houses and lighthouse in the distance.		Dog Park	Canis Beach	General	http://maps.secondlife.com/secondlife/Canis%20Beach/71/61/21	VE	0	2
89	18	5	Medieval stone houses, fitted with simple, minimal furniture, all near a forest		The Alchemist	Alice	General	http://maps.secondlife.com/secondlife/Alice/99/17/83	VE	0	2
90	18	5	Large natural park, with trees, tall grass and bushes in autumn colours		Alirium	Alirium	General	http://maps.secondlife.com/secondlife/alirium/145/129/1607	VE	0	2
91	19	5	Aquarium tanks on display in large advertising hall		Aquarium	Trianwe	Moderate	http://maps.secondlife.com/secondlife/Trianwe/194/13/511	VE	0	1
92	19	5	Ancient Roman/Greek-inspired town, with stone roads, walls, temples, and a small port		City of Tentium	Calypso Reef	Adult	http://maps.secondlife.com/secondlife/Calypso%20Reef/34/17/23	VE	0	1
93	19	5	Large sports complex, with basketball court		TT	TT Enterprises	Moderate	http://maps.secondlife.com/secondlife/TT%20Enterprises/180/115/26	VE	0	1
94	19	5	Air base with stationed aircraft		Salon de Provence	BA 701	Moderate	http://maps.secondlife.com/secondlife/BA%20701/211/173/24	VE	0	1

No	Block	Cluster	ResearcherDescription	Second LifeURLTag	Parcel	Region	Rating	URL	Type	Music	Session
95	19	5	Rainforest with palm trees and lush vegetation, close to a sandy beach and some huts		Amazon Rainforest	Alexandre	Moderate	http://maps.secondlife.com/secondlife/Alexandre/207/102/25	VE	0	1
96	20	5	Landscape with a swamp, with plenty of stilt houses, shacks, trees and reeds		Noctis Swamp Shack	New Tolouse Bayou	Moderate	http://maps.secondlife.com/secondlife/New%20Toulouse%20Bayou/57/163/27	VE	0	1
97	20	5	Large medical facility with receptions, examination and waiting rooms, and multiple floors		Second Health	Second Health London	General	http://maps.secondlife.com/secondlife/Second%20Health%20London/79/228/27	VE	0	1
98	20	5	Military exhibition with weapons, tanks, stationed aircraft		Equino's Armory	Datana	Moderate	http://maps.secondlife.com/secondlife/Datana/237/189/54	VE	0	1
99	20	5	Small railway station facing the sea on one side, and mountains on the other. Train tracks present, extending to the left and right.		Great Second Life Railway	Slate	General	http://maps.secondlife.com/secondlife/Slate/134/214/42	VE	0	1
100	20	5	Treehouse in a jungle, looking out on a river, trees, bushes and vines, and rocks		Makeahla Jungle	Makeahla	Moderate	http://maps.secondlife.com/secondlife/Makeahla/90/150/36	VE	0	1

E.2 Participant instruction sheet

INFORMATION SHEET: Emotional ratings experiment

http://pat-harding.com/profile-data/wp-content/uploads/2014/05/Oculus-Rift-DK1-CAD-Test.jpg

In this experiment, you will first respond to some questionnaires and then be shown several virtual environments. After each environment, you will be asked to rate how *you would feel in real life if you experienced what it depicts*. Please do not rate the stimulus in itself, but rather how *you would feel if interacting with it*.

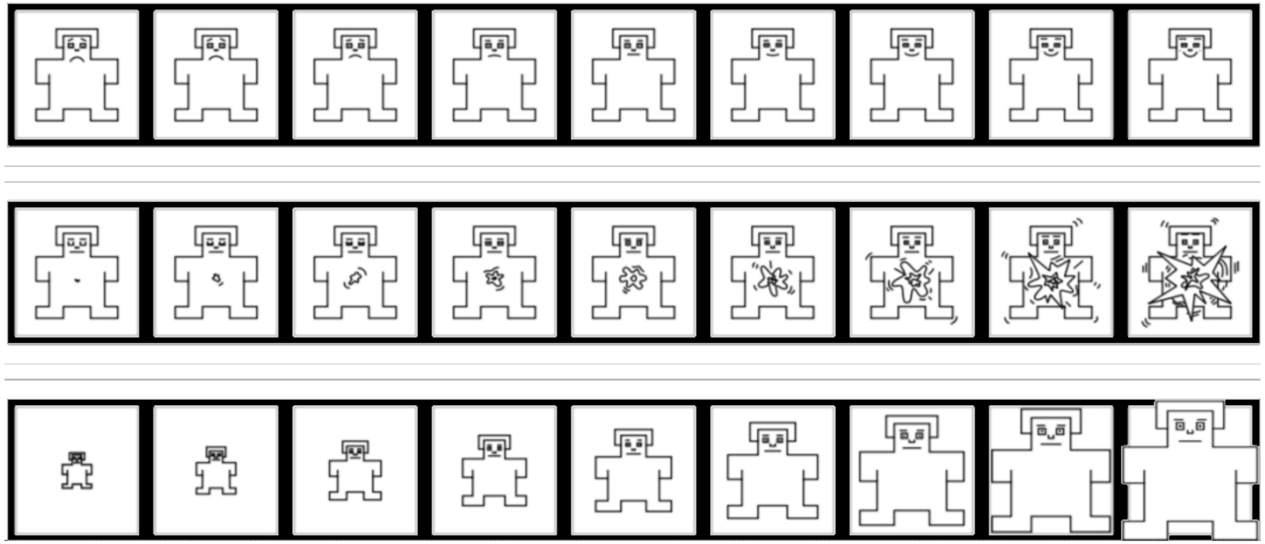
For displaying the virtual environments, we will be using `Second Life` and a `Head Mounted Display` (image to the left).

`Second Life` is an online collection of virtual worlds, between which you can `teleport`. `Second Life` is not meant to be an online game, but rather a place for people (appearing as avatars) to socialise and explore surroundings. However, you should be aware your task here is not to socialise, but *simply to freely explore* these virtual worlds and rate how they make you feel.

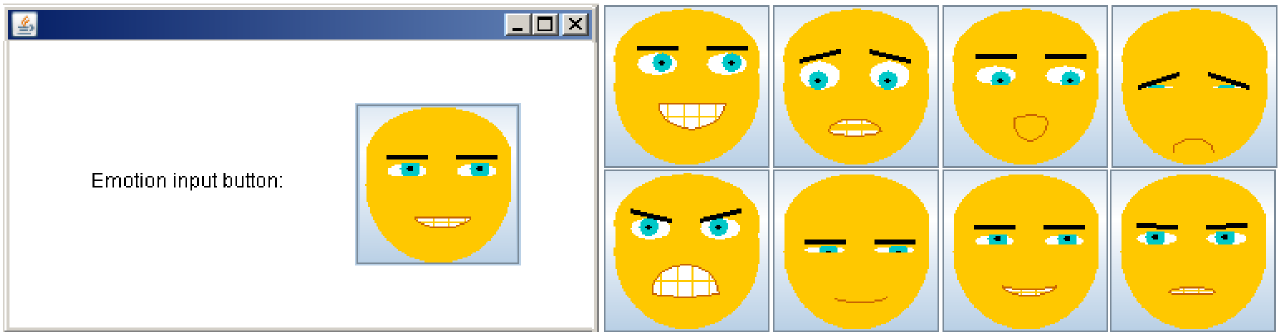
We will first practise how to put on the `Head Mounted Display`, how to move around / open doors in these virtual worlds, and how to provide responses. The `Head Mounted Display` will enable you to see the virtual environments as if you were part of them. For instance, you can move your head and the image will change accordingly to let you see in the direction you turned towards.

The rating procedure for these environments will involve **<u>two types of responses</u>**:

<u>1.</u> Below you have 3 sets of images, the first going from sad to happy, the second from bored/relaxed to alert, and the third from submissive to dominant. From each set you will have to pick one image which best describes how you would feel in a situation related to the stimulus shown.



2. Below you have a cartoon figure which will change expression depending on how you move the mouse over it. When you think you have found an expression matching how you feel, make a click on that area of the face to record your response.



PARTICIPANTS' RIGHTS
Please note that you are free to stop the experiment at any point if you so wish. If you have any questions as a result of reading this information sheet, you should ask the researcher before the study begins.

TIME COMMITMENT & REIMBURSEMENT
This study will comprise of a session of roughly 1h. You will receive £7 in return for your participation. Payment will be given at the end of the experiment.

CONFIDENTIALITY/ANONYMITY
The data we collect do not contain any personally identifiable information (name or initials & email for raffle) about you except on the Informed Consent form which will follow. Your responses within the experiment will not be linked with the identifying information you supplied (name or initials & email).

FOR FURTHER INFORMATION
Caterina Constantinescu - the researcher - will be glad to answer your questions about this study at any time. You may contact her at caterina.constantinescu@ed.ac.uk. Alternatively, the staff members supervising this project can also be reached: Dr. Sarah E. MacPherson (sarah.macpherson@ed.ac.uk) and Dr. Adam Moore (amoore23@exseed.ed.ac.uk).

Thank you for your participation!

INFORMED CONSENT FORM

By signing below, you agreeing that:
(1) you have read and understood the Information Sheet,
(2) any questions about your participation in this study have been answered satisfactorily,
(3) you are taking part in this research study voluntarily (without coercion),
(4) you understand that participation in this study involves completion of some standardised tests [Toronto Alexithymia Scale-20, Patient Health Questionnaire-2] which are used as preliminary screens for clinical conditions of which you may not be aware. Scores from these tests would not be sufficient basis for clinical decisions or diagnosis, contain substantial margins of error, and are not used for diagnostic purposes in this study. Though it is not possible to provide feedback of individual scores to participants, these scores might hint at health problems. In the event that I produce scores of potential clinical concern, researchers should (check one and provide relevant contact information):

☐ Contact me at:______________________________________

☐ Contact my GP at: ___________________________________

☐ Do nothing. I absolve the researchers of any obligation to contact me about this.

Participant's name:*	Email:	Signature:	Date:

** Participants may use their initials if they wish.*

Person obtaining consent:	Signature:
Caterina Constantinescu (caterina.constantinescu@ed.ac.uk)	

Supervisory team for study: Dr. Sarah E. MacPherson (sarah.macpherson@ed.ac.uk),
Dr. Adam Moore (amoore23@exseed.ed.ac.uk).

E.3 Selection of 25 virtual environment stimuli

Sheet 001 of 002

Agriopis.png

Alice.png

AngelManor.png

Bracket.png

BrandyWineIsland.png

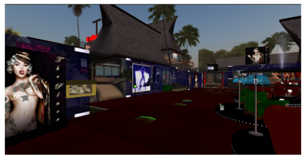
Brothel.png

CanisBeach.png

Cavettaz.png

DepozSpecialities.png

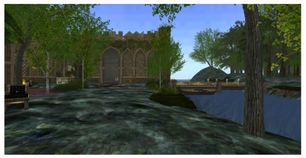
EscortOasis.png



Guerreiros.png

Harshap.png

Kalepa.png



KismetNorthwinds.png

MediterraneoOC.png

Sheet 002 of 002

Misthaven.png

MountainsOfCreta.png

Picklemoon.png

Picklesong.png

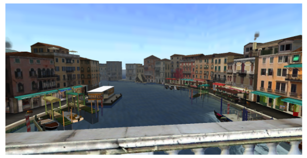
PortBabbage.png

Prefabrica.png

QuietlyTuesday.png

Trianwe.png

Triglav.png



YumixPrada.png

E.4 R session information

```
R version 3.4.0 (2017-04-21)
Platform: x86_64-pc-linux-gnu (64-bit)
Running under: Ubuntu 16.04.2 LTS

Matrix products: default
BLAS: /usr/lib/libblas/libblas.so.3.6.0
LAPACK: /usr/lib/lapack/liblapack.so.3.6.0

locale:
 [1] LC_CTYPE=en_GB.UTF-8       LC_NUMERIC=C               LC_TIME=en_GB.UTF-8        LC_COLLATE=en_GB.UTF-8
 [5] LC_MONETARY=en_GB.UTF-8    LC_MESSAGES=en_GB.UTF-8    LC_PAPER=en_GB.UTF-8       LC_NAME=C
 [9] LC_ADDRESS=C               LC_TELEPHONE=C             LC_MEASUREMENT=en_GB.UTF-8 LC_IDENTIFICATION=C

attached base packages:
[1] parallel  stats     graphics  grDevices utils     datasets  methods   base

other attached packages:
 [1] psych_1.7.5                  paran_1.5.1                  MASS_7.3-47                  xtable_1.8-2
 [5] texreg_1.36.23               LMERConvenienceFunctions_2.10 lme4_1.1-13                 Matrix_1.2-10
 [9] clusteval_0.1                mclust_5.3                   robustHD_0.5.1               perry_0.2.0
[13] robustbase_0.92-7            broom_0.4.2                  R.utils_2.5.0                R.oo_1.21.0
[17] R.methodsS3_1.7.1            matrixStats_0.52.2           purrr_0.2.2.2                tidyr_0.6.3
[21] plyr_1.8.4                   data.table_1.10.4            stringr_1.2.0                gridExtra_2.2.1
[25] wordcloud_2.5                RColorBrewer_1.1-2           rgl_0.98.1                   scales_0.4.1
[29] ggrepel_0.6.5                directlabels_2017.03.31      ggplot2_2.2.1

loaded via a namespace (and not attached):
```

```
 [1] Rcpp_0.12.10     mvtnorm_1.0-6    lattice_0.20-35  assertthat_0.2.0 digest_0.6.12    mime_0.5         slam_0.1-40
 [8] R6_2.2.1         spam_2.1-1       rlang_0.1        lazyeval_0.2.0   minqa_1.2.4      nloptr_1.0.4     splines_3.4.0
[15] foreign_0.8-67   htmlwidgets_0.9  munsell_0.4.3    shiny_1.0.3      compiler_3.4.0   httpuv_1.3.5     pkgconfig_2.0.1
[22] mnormt_1.5-5     mgcv_1.8-17      htmltools_0.3.6  tibble_1.3.1     quadprog_1.5-5   dplyr_0.7.2      grid_3.4.0
[29] nlme_3.1-131     jsonlite_1.4     gtable_0.2.0     magrittr_1.5     stringi_1.1.5    reshape2_1.4.2   bindrcpp_0.2
[36] tools_3.4.0      LCFdata_2.0      glue_1.1.1       DEoptimR_1.0-8   maps_3.2.0       fields_9.0       colorspace_1.3-2
[43] dotCall64_0.9-04 knitr_1.16       bindr_0.1
```

Appendix F

Discrepancies between emotional data as classified by cluster analysis algorithms vs. human participants

F.1 Participant instruction sheet

INFORMATION SHEET: Sorting emotional virtual environments

In this experiment, you will first be asked to respond to some QUESTIONNAIRES, and then view a series of VIRTUAL ENVIRONMENTS. After viewing all the environments, you will be asked to:

- Sort them into groups,
- Rate them according to several scales (which will be described during a practice phase, shortly),

depending on how you would feel in real life if you experienced what they depict. On a random basis, you might be asked to start with the sorting task, or rather, with the rating task.

FOR THE SORTING TASK, you will be able to create **AS MANY CATEGORIES OF STIMULI AS YOU LIKE**. Within each category, you can also **ORDER THE STIMULI FROM THE BEST (POSITION 1) TO THE LEAST GOOD REPRESENTATIVE OF THE CATEGORY (LAST POSITION)**. Please be aware that a category of **'UNCLASSIFIABLE'** content will also be provided for you from the start of the task, but you should only use this if you think it is necessary (i.e., you can leave this category empty or not, depending on how well you think the stimuli fit within the other categories you have created).

FOR THE RATING TASK, you will be given further instructions during the practice.

ADDITIONAL INFORMATION

For the virtual environments, we will be using Second Life software and a Head Mounted Display, **which you will need to put on to view each environment, and then take off to sort the environments, and give some emotional ratings.**

Second Life is an online collection of virtual worlds, between which you can teleport freely. Second Life is not meant to be an online game, but rather a place for people (appearing as avatars) to socialise and explore surroundings. However, you should be aware your task is not to socialise, but ***simply to freely explore*** **these virtual worlds during the time given. When the exploration time is up, you will hear an alarm and then be able to view the next environment.**

The head mounted display is called the Oculus Rift (image below) and it enables you to see the virtual environments as if you were part of them. For instance, you can move your head and the image will change accordingly to let you see in the direction you turned towards. **To walk in Second Life while having the Oculus Rift on, you can use the direction keys: ← and → . You can also look around by moving your head in real life, at the same time.**

In the virtual environments condition, we will first practise how to put on the head mounted display, how to move around and open doors in these virtual worlds and how to provide ratings and sort the stimuli. **NOTE: USING HEAD MOUNTED DISPLAYS OVER A LONG PERIOD CAN LEAD TO SLIGHT FEELINGS OF MOTION SICKNESS, SO BE SURE TO ASK FOR A SHORT BREAK IF YOU FEEL YOU NEED ONE!**

F.2 Assessing similarities, by region (VE) and reference date

No	Region	ReferenceScan	2016-10-10	2016-10-12	2016-10-14	2016-10-17	2016-10-19	2016-10-21	2016-10-24	2016-10-26	2016-10-28	2016-10-31
1	Agriopis	2016-10-10	1	0.748	0.835	0.841	0.861	0.828	0.847	0.834	0.837	0.844
2	Agriopis	2016-10-12	0.748	1	0.825	0.796	0.807	0.845	0.793	0.817	0.822	0.795
3	Agriopis	2016-10-14	0.835	0.825	1.001	0.918	0.909	0.922	0.913	0.941	0.943	0.917
4	Agriopis	2016-10-17	0.841	0.796	0.918	1.001	0.939	0.906	0.972	0.939	0.936	0.974
5	Agriopis	2016-10-19	0.861	0.807	0.909	0.939	1	0.932	0.945	0.93	0.924	0.933
6	Agriopis	2016-10-21	0.828	0.845	0.922	0.906	0.932	1	0.911	0.938	0.937	0.903
7	Agriopis	2016-10-24	0.847	0.793	0.913	0.972	0.945	0.911	1.001	0.934	0.93	0.968
8	Agriopis	2016-10-26	0.834	0.817	0.941	0.939	0.93	0.938	0.934	1.001	0.976	0.932
9	Agriopis	2016-10-28	0.837	0.822	0.943	0.936	0.924	0.937	0.93	0.976	1.001	0.938
10	Agriopis	2016-10-31	0.844	0.795	0.917	0.974	0.933	0.903	0.968	0.932	0.938	1.001
11	Alice	2016-10-10	1	0.9	0.926	0.91	0.885	0.907	0.912	0.878	0.888	0.906
12	Alice	2016-10-12	0.9	1.004	0.952	0.936	0.952	0.937	0.933	0.953	0.945	0.943
13	Alice	2016-10-14	0.926	0.952	1	0.962	0.924	0.971	0.962	0.928	0.949	0.968
14	Alice	2016-10-17	0.91	0.936	0.962	1	0.942	0.967	0.97	0.929	0.934	0.959
15	Alice	2016-10-19	0.885	0.952	0.924	0.942	1	0.94	0.94	0.972	0.95	0.933
16	Alice	2016-10-21	0.907	0.937	0.971	0.967	0.94	1	0.97	0.932	0.954	0.975
17	Alice	2016-10-24	0.912	0.933	0.962	0.97	0.94	0.97	1	0.925	0.936	0.956
18	Alice	2016-10-26	0.878	0.953	0.928	0.929	0.972	0.932	0.925	1	0.969	0.938
19	Alice	2016-10-28	0.888	0.945	0.949	0.934	0.95	0.954	0.936	0.969	1	0.951
20	Alice	2016-10-31	0.906	0.943	0.968	0.959	0.933	0.975	0.956	0.938	0.951	1
21	Angel Manor	2016-10-10	1.003	0.905	0.905	0.899	0.888	0.881	0.877	0.824	0.892	0.888
22	Angel Manor	2016-10-12	0.905	1.004	0.992	0.973	0.971	0.963	0.96	0.858	0.95	0.952
23	Angel Manor	2016-10-14	0.905	0.992	1.004	0.971	0.966	0.962	0.956	0.857	0.953	0.953
24	Angel Manor	2016-10-17	0.899	0.973	0.971	1.004	0.976	0.974	0.966	0.858	0.963	0.963
25	Angel Manor	2016-10-19	0.888	0.971	0.966	0.976	1.022	0.989	0.983	0.876	0.982	0.972
26	Angel Manor	2016-10-21	0.881	0.963	0.962	0.974	0.989	1.004	0.984	0.861	0.975	0.977
27	Angel Manor	2016-10-24	0.877	0.96	0.956	0.966	0.983	0.984	1.004	0.862	0.972	0.98
28	Angel Manor	2016-10-26	0.824	0.858	0.857	0.858	0.876	0.861	0.862	1.007	0.845	0.849
29	Angel Manor	2016-10-28	0.892	0.95	0.953	0.963	0.982	0.975	0.972	0.845	1.003	0.975
30	Angel Manor	2016-10-31	0.888	0.952	0.953	0.963	0.972	0.977	0.98	0.849	0.975	1.003
31	Bracket	2016-10-10	1.002	0.916	0.606	0.913	0.909	0.886	0.865	0.885	0.893	0.877
32	Bracket	2016-10-12	0.916	1.002	0.599	0.989	0.953	0.939	0.901	0.935	0.938	0.917
33	Bracket	2016-10-14	0.606	0.599	1.001	0.6	0.631	0.643	0.668	0.645	0.628	0.642
34	Bracket	2016-10-17	0.913	0.989	0.6	1.002	0.955	0.94	0.901	0.937	0.94	0.92
35	Bracket	2016-10-19	0.909	0.953	0.631	0.955	1.002	0.929	0.911	0.932	0.954	0.946
36	Bracket	2016-10-21	0.886	0.939	0.643	0.94	0.929	1.002	0.866	0.971	0.9	0.925
37	Bracket	2016-10-24	0.865	0.901	0.668	0.901	0.911	0.866	1.002	0.872	0.927	0.898
38	Bracket	2016-10-26	0.885	0.935	0.645	0.937	0.932	0.971	0.872	1.002	0.906	0.924
39	Bracket	2016-10-28	0.893	0.938	0.628	0.94	0.954	0.9	0.927	0.906	1.002	0.941
40	Bracket	2016-10-31	0.877	0.917	0.642	0.92	0.946	0.925	0.898	0.924	0.941	1.002
41	Brandy Wine Island	2016-10-10	1.002	0.756	0.013	0.004	0.004	0.004	0.004	0.004	0.004	0.004
42	Brandy Wine Island	2016-10-12	0.756	1.004	0.015	0.007	0.007	0.007	0.007	0.007	0.007	0.007
43	Brandy Wine Island	2016-10-14	0.013	0.015	1.011	0.022	0.022	0.022	0.022	0.022	0.022	0.022
44	Brandy Wine Island	2016-10-17	0.004	0.007	0.022	NA	NA	NA	NA	NA	NA	NA
45	Brandy Wine Island	2016-10-19	0.004	0.007	0.022	NA	NA	NA	NA	NA	NA	NA
46	Brandy Wine Island	2016-10-21	0.004	0.007	0.022	NA	NA	NA	NA	NA	NA	NA
47	Brandy Wine Island	2016-10-24	0.004	0.007	0.022	NA	NA	NA	NA	NA	NA	NA
48	Brandy Wine Island	2016-10-26	0.004	0.007	0.022	NA	NA	NA	NA	NA	NA	NA
49	Brandy Wine Island	2016-10-28	0.004	0.007	0.022	NA	NA	NA	NA	NA	NA	NA
50	Brandy Wine Island	2016-10-31	0.004	0.007	0.022	NA	NA	NA	NA	NA	NA	NA
51	BROTHEL	2016-10-10	1.001	0.794	0.803	0.786	0.791	0.786	0.789	0.776	0.79	0.79

No	Region	ReferenceScan	2016-10-10	2016-10-12	2016-10-14	2016-10-17	2016-10-19	2016-10-21	2016-10-24	2016-10-26	2016-10-28	2016-10-31
52	BROTHEL	2016-10-12	0.794	1.003	0.78	0.759	0.761	0.751	0.75	0.742	0.772	0.765
53	BROTHEL	2016-10-14	0.803	0.78	1.003	0.77	0.778	0.767	0.774	0.751	0.783	0.779
54	BROTHEL	2016-10-17	0.786	0.759	0.77	1.003	0.776	0.796	0.769	0.759	0.78	0.769
55	BROTHEL	2016-10-19	0.791	0.761	0.778	0.776	1.001	0.766	0.785	0.772	0.788	0.785
56	BROTHEL	2016-10-21	0.786	0.751	0.767	0.796	0.766	1.001	0.788	0.772	0.791	0.773
57	BROTHEL	2016-10-24	0.789	0.75	0.774	0.769	0.785	0.788	1.003	0.806	0.799	0.783
58	BROTHEL	2016-10-26	0.776	0.742	0.751	0.759	0.772	0.772	0.806	1.002	0.788	0.775
59	BROTHEL	2016-10-28	0.79	0.772	0.783	0.78	0.788	0.791	0.799	0.788	1.004	0.798
60	BROTHEL	2016-10-31	0.79	0.765	0.779	0.769	0.785	0.773	0.783	0.775	0.798	1.002
61	Canis Beach	2016-10-10	1.003	0.665	0.67	0.68	0.678	0.679	0.684	0.663	0.669	0.673
62	Canis Beach	2016-10-12	0.665	1	0.786	0.856	0.85	0.839	0.816	0.833	0.84	0.848
63	Canis Beach	2016-10-14	0.67	0.786	1.006	0.809	0.81	0.817	0.832	0.802	0.822	0.807
64	Canis Beach	2016-10-17	0.68	0.856	0.809	1.003	0.876	0.87	0.842	0.824	0.846	0.857
65	Canis Beach	2016-10-19	0.678	0.85	0.81	0.876	1.004	0.861	0.847	0.832	0.853	0.866
66	Canis Beach	2016-10-21	0.679	0.839	0.817	0.87	0.861	1.003	0.852	0.834	0.861	0.839
67	Canis Beach	2016-10-24	0.684	0.816	0.832	0.842	0.847	0.852	1.004	0.831	0.857	0.848
68	Canis Beach	2016-10-26	0.663	0.833	0.802	0.824	0.832	0.834	0.831	1.003	0.853	0.842
69	Canis Beach	2016-10-28	0.669	0.84	0.822	0.846	0.853	0.861	0.857	0.853	1.002	0.857
70	Canis Beach	2016-10-31	0.673	0.848	0.807	0.857	0.866	0.839	0.848	0.842	0.857	1.004
71	Cavettaz	2016-10-10	1	0.856	0.996	0.993	0.993	0.993	0.996	0.996	0.996	0.994
72	Cavettaz	2016-10-12	0.856	1	0.857	0.855	0.858	0.854	0.857	0.857	0.857	0.854
73	Cavettaz	2016-10-14	0.996	0.857	1	0.995	0.995	0.994	0.997	0.997	0.998	0.995
74	Cavettaz	2016-10-17	0.993	0.855	0.995	1	0.992	0.992	0.995	0.995	0.995	0.992
75	Cavettaz	2016-10-19	0.993	0.858	0.995	0.992	1	0.991	0.994	0.994	0.995	0.992
76	Cavettaz	2016-10-21	0.993	0.854	0.994	0.992	0.991	1	0.994	0.994	0.994	0.992
77	Cavettaz	2016-10-24	0.996	0.857	0.997	0.995	0.994	0.994	1	0.998	0.997	0.994
78	Cavettaz	2016-10-26	0.996	0.857	0.997	0.995	0.994	0.994	0.998	1	0.997	0.994
79	Cavettaz	2016-10-28	0.996	0.857	0.998	0.995	0.995	0.994	0.997	0.997	1	0.995
80	Cavettaz	2016-10-31	0.994	0.854	0.995	0.992	0.992	0.992	0.994	0.994	0.995	1
81	Depoz Specialties	2016-10-10	1	0.96	0.958	0.957	0.961	0.958	0.957	0.96	0.959	0.96
82	Depoz Specialties	2016-10-12	0.96	1	0.995	0.997	0.988	0.993	0.994	0.989	0.993	0.993
83	Depoz Specialties	2016-10-14	0.958	0.995	1	0.997	0.988	0.992	0.994	0.989	0.993	0.994
84	Depoz Specialties	2016-10-17	0.957	0.997	0.997	1	0.991	0.994	0.995	0.99	0.993	0.994
85	Depoz Specialties	2016-10-19	0.961	0.988	0.988	0.991	1	0.987	0.988	0.994	0.987	0.987
86	Depoz Specialties	2016-10-21	0.958	0.993	0.992	0.994	0.987	1	0.997	0.989	0.995	0.995
87	Depoz Specialties	2016-10-24	0.957	0.994	0.994	0.995	0.988	0.997	1	0.99	0.994	0.995
88	Depoz Specialties	2016-10-26	0.96	0.989	0.989	0.99	0.994	0.989	0.99	1	0.99	0.99
89	Depoz Specialties	2016-10-28	0.959	0.993	0.993	0.993	0.987	0.995	0.994	0.99	1	0.996
90	Depoz Specialties	2016-10-31	0.96	0.993	0.994	0.994	0.987	0.995	0.995	0.99	0.996	1
91	escort oasis	2016-10-10	1.007	0.882	0.968	0.946	0.955	0.961	0.911	0.951	0.965	0.874
92	escort oasis	2016-10-12	0.882	1.006	0.907	0.894	0.892	0.891	0.922	0.888	0.901	0.925
93	escort oasis	2016-10-14	0.968	0.907	1.007	0.961	0.97	0.971	0.934	0.965	0.981	0.897
94	escort oasis	2016-10-17	0.946	0.894	0.961	1.006	0.951	0.959	0.912	0.951	0.966	0.884
95	escort oasis	2016-10-19	0.955	0.892	0.97	0.951	1.006	0.968	0.925	0.958	0.979	0.886
96	escort oasis	2016-10-21	0.961	0.891	0.971	0.959	0.968	1.006	0.926	0.962	0.982	0.887
97	escort oasis	2016-10-24	0.911	0.922	0.934	0.912	0.925	0.926	1.005	0.918	0.939	0.911
98	escort oasis	2016-10-26	0.951	0.888	0.965	0.951	0.958	0.962	0.918	1.006	0.977	0.886
99	escort oasis	2016-10-28	0.965	0.901	0.981	0.966	0.979	0.982	0.939	0.977	1.007	0.899
100	escort oasis	2016-10-31	0.874	0.925	0.897	0.884	0.886	0.887	0.911	0.886	0.899	1.005
101	GUERREIROS	2016-10-10	1.003	0.87	0.838	0.839	0.835	0.838	0.819	0.831	0.82	0.771
102	GUERREIROS	2016-10-12	0.87	1.004	0.953	0.948	0.954	0.946	0.925	0.941	0.92	0.867
103	GUERREIROS	2016-10-14	0.838	0.953	1.003	0.946	0.942	0.938	0.916	0.93	0.913	0.857

No	Region	ReferenceScan	2016-10-10	2016-10-12	2016-10-14	2016-10-17	2016-10-19	2016-10-21	2016-10-24	2016-10-26	2016-10-28	2016-10-31
104	GUERREIROS	2016-10-17	0.839	0.948	0.946	1.003	0.95	0.949	0.927	0.941	0.922	0.885
105	GUERREIROS	2016-10-19	0.835	0.954	0.942	0.95	1.003	0.953	0.934	0.948	0.924	0.871
106	GUERREIROS	2016-10-21	0.838	0.946	0.938	0.949	0.953	1.003	0.967	0.951	0.937	0.878
107	GUERREIROS	2016-10-24	0.819	0.925	0.916	0.927	0.934	0.967	1.003	0.942	0.922	0.869
108	GUERREIROS	2016-10-26	0.831	0.941	0.93	0.941	0.948	0.951	0.942	1.003	0.976	0.897
109	GUERREIROS	2016-10-28	0.82	0.92	0.913	0.922	0.924	0.937	0.922	0.976	1.003	0.882
110	GUERREIROS	2016-10-31	0.771	0.867	0.857	0.885	0.871	0.878	0.869	0.897	0.882	1.003
111	Harshap	2016-10-10	1.002	0.946	0.955	0.951	0.954	0.949	0.937	0.945	0.947	0.953
112	Harshap	2016-10-12	0.946	1.002	0.969	0.962	0.958	0.956	0.945	0.938	0.958	0.956
113	Harshap	2016-10-14	0.955	0.969	1.002	0.977	0.974	0.977	0.964	0.949	0.974	0.977
114	Harshap	2016-10-17	0.951	0.962	0.977	1.002	0.979	0.982	0.967	0.953	0.978	0.978
115	Harshap	2016-10-19	0.954	0.958	0.974	0.979	1.002	0.973	0.961	0.951	0.975	0.978
116	Harshap	2016-10-21	0.949	0.956	0.977	0.982	0.973	1.002	0.974	0.961	0.982	0.986
117	Harshap	2016-10-24	0.937	0.945	0.964	0.967	0.961	0.974	1.002	0.951	0.968	0.978
118	Harshap	2016-10-26	0.945	0.938	0.949	0.953	0.951	0.961	0.951	1.002	0.956	0.96
119	Harshap	2016-10-28	0.947	0.958	0.974	0.978	0.975	0.982	0.968	0.956	1.002	0.984
120	Harshap	2016-10-31	0.953	0.956	0.977	0.978	0.978	0.986	0.978	0.96	0.984	1.002
121	Kalepa	2016-10-10	1	0.824	0.802	0.819	0.812	0.79	0.795	0.808	0.809	0.814
122	Kalepa	2016-10-12	0.824	1	0.925	0.93	0.921	0.91	0.909	0.913	0.924	0.937
123	Kalepa	2016-10-14	0.802	0.925	1	0.934	0.901	0.893	0.912	0.886	0.901	0.919
124	Kalepa	2016-10-17	0.819	0.93	0.934	1	0.907	0.902	0.909	0.887	0.9	0.919
125	Kalepa	2016-10-19	0.812	0.921	0.901	0.907	1	0.897	0.893	0.895	0.899	0.915
126	Kalepa	2016-10-21	0.79	0.91	0.893	0.902	0.897	1.002	0.868	0.873	0.886	0.904
127	Kalepa	2016-10-24	0.795	0.909	0.912	0.909	0.893	0.868	1	0.886	0.89	0.905
128	Kalepa	2016-10-26	0.808	0.913	0.886	0.887	0.895	0.873	0.886	1	0.901	0.914
129	Kalepa	2016-10-28	0.809	0.924	0.901	0.9	0.899	0.886	0.89	0.901	1	0.922
130	Kalepa	2016-10-31	0.814	0.937	0.919	0.919	0.915	0.904	0.905	0.914	0.922	1
131	Kismet Northwinds	2016-10-10	1.002	0.855	0.915	0.919	0.919	0.928	0.922	0.928	0.923	0.922
132	Kismet Northwinds	2016-10-12	0.855	1.003	0.856	0.875	0.866	0.874	0.858	0.868	0.867	0.866
133	Kismet Northwinds	2016-10-14	0.915	0.856	1.002	0.954	0.952	0.963	0.955	0.963	0.962	0.955
134	Kismet Northwinds	2016-10-17	0.919	0.875	0.954	1.002	0.973	0.972	0.958	0.972	0.967	0.964
135	Kismet Northwinds	2016-10-19	0.919	0.866	0.952	0.973	1.002	0.971	0.966	0.968	0.963	0.964
136	Kismet Northwinds	2016-10-21	0.928	0.874	0.963	0.972	0.971	1.002	0.968	0.981	0.98	0.98
137	Kismet Northwinds	2016-10-24	0.922	0.858	0.955	0.958	0.966	0.968	1.002	0.972	0.973	0.972
138	Kismet Northwinds	2016-10-26	0.928	0.868	0.963	0.972	0.968	0.981	0.972	1.002	0.98	0.98
139	Kismet Northwinds	2016-10-28	0.923	0.867	0.962	0.967	0.963	0.98	0.973	0.98	1.002	0.978
140	Kismet Northwinds	2016-10-31	0.922	0.866	0.955	0.964	0.964	0.98	0.972	0.98	0.978	1.002
141	Mediterraneo OC	2016-10-10	1.001	0.64	0.66	0.697	0.648	0.711	0.647	0.637	0.639	0.64
142	Mediterraneo OC	2016-10-12	0.64	1.015	0.706	0.672	0.693	0.712	0.666	0.659	0.662	0.684
143	Mediterraneo OC	2016-10-14	0.66	0.706	1.001	0.714	0.745	0.714	0.686	0.729	0.688	0.732
144	Mediterraneo OC	2016-10-17	0.697	0.672	0.714	1.003	0.723	0.742	0.75	0.692	0.727	0.678
145	Mediterraneo OC	2016-10-19	0.648	0.693	0.745	0.723	1.003	0.712	0.711	0.721	0.708	0.688
146	Mediterraneo OC	2016-10-21	0.711	0.712	0.714	0.742	0.712	1.002	0.701	0.714	0.7	0.712
147	Mediterraneo OC	2016-10-24	0.647	0.666	0.686	0.75	0.711	0.701	1.003	0.703	0.682	0.677
148	Mediterraneo OC	2016-10-26	0.637	0.659	0.729	0.692	0.721	0.714	0.703	1.001	0.702	0.723
149	Mediterraneo OC	2016-10-28	0.639	0.662	0.688	0.727	0.708	0.7	0.682	0.702	1.002	0.679
150	Mediterraneo OC	2016-10-31	0.64	0.684	0.732	0.678	0.688	0.712	0.677	0.723	0.679	1.001
151	Misthaven	2016-10-10	1.001	0.924	0.943	0.908	0.854	0.854	0.853	0.853	0.851	0.847
152	Misthaven	2016-10-12	0.924	1.001	0.948	0.93	0.871	0.865	0.868	0.866	0.865	0.862
153	Misthaven	2016-10-14	0.943	0.948	1.001	0.939	0.887	0.886	0.885	0.891	0.887	0.884
154	Misthaven	2016-10-17	0.908	0.93	0.939	1.001	0.903	0.892	0.891	0.879	0.88	0.882
155	Misthaven	2016-10-19	0.854	0.871	0.887	0.903	1.001	0.978	0.965	0.964	0.956	0.959

No	Region	ReferenceScan	2016-10-10	2016-10-12	2016-10-14	2016-10-17	2016-10-19	2016-10-21	2016-10-24	2016-10-26	2016-10-28	2016-10-31
156	Misthaven	2016-10-21	0.854	0.865	0.886	0.892	0.978	1.001	0.965	0.965	0.951	0.965
157	Misthaven	2016-10-24	0.853	0.868	0.885	0.891	0.965	0.965	1.001	0.968	0.951	0.965
158	Misthaven	2016-10-26	0.853	0.866	0.891	0.879	0.964	0.965	0.968	1.001	0.966	0.965
159	Misthaven	2016-10-28	0.851	0.865	0.887	0.88	0.956	0.951	0.951	0.966	1.001	0.967
160	Misthaven	2016-10-31	0.847	0.862	0.884	0.882	0.959	0.965	0.965	0.965	0.967	1.001
161	Mountains of Creta	2016-10-10	1	1.065	0.916	0.885	0.913	0.914	0.905	0.885	0.911	0.909
162	Mountains of Creta	2016-10-12	1.065	1.304	1.1	1.07	1.09	1.099	1.085	1.072	1.096	1.1
163	Mountains of Creta	2016-10-14	0.916	1.1	1	0.926	0.949	0.959	0.942	0.929	0.957	0.955
164	Mountains of Creta	2016-10-17	0.885	1.07	0.926	1	0.925	0.93	0.907	0.904	0.93	0.928
165	Mountains of Creta	2016-10-19	0.913	1.09	0.949	0.925	1	0.954	0.93	0.926	0.953	0.951
166	Mountains of Creta	2016-10-21	0.914	1.099	0.959	0.93	0.954	1	0.94	0.936	0.959	0.957
167	Mountains of Creta	2016-10-24	0.905	1.085	0.942	0.907	0.93	0.94	1	0.908	0.937	0.933
168	Mountains of Creta	2016-10-26	0.885	1.072	0.929	0.904	0.926	0.936	0.908	1	0.935	0.931
169	Mountains of Creta	2016-10-28	0.911	1.096	0.957	0.93	0.953	0.959	0.937	0.935	1	0.961
170	Mountains of Creta	2016-10-31	0.909	1.1	0.955	0.928	0.951	0.957	0.933	0.931	0.961	1
171	Picklemoon	2016-10-10	1	0.946	0.972	0.949	0.962	0.96	0.954	0.949	0.952	0.928
172	Picklemoon	2016-10-12	0.946	1	0.946	0.985	0.95	0.957	0.952	0.966	0.949	0.928
173	Picklemoon	2016-10-14	0.972	0.946	1	0.949	0.979	0.972	0.968	0.948	0.964	0.943
174	Picklemoon	2016-10-17	0.949	0.985	0.949	1	0.952	0.959	0.954	0.969	0.944	0.931
175	Picklemoon	2016-10-19	0.962	0.95	0.979	0.952	1	0.981	0.97	0.952	0.972	0.948
176	Picklemoon	2016-10-21	0.96	0.957	0.972	0.959	0.981	1	0.961	0.952	0.973	0.94
177	Picklemoon	2016-10-24	0.954	0.952	0.968	0.954	0.97	0.961	1	0.964	0.966	0.956
178	Picklemoon	2016-10-26	0.949	0.966	0.948	0.969	0.952	0.952	0.964	1	0.958	0.938
179	Picklemoon	2016-10-28	0.952	0.949	0.964	0.944	0.972	0.973	0.966	0.958	1	0.946
180	Picklemoon	2016-10-31	0.928	0.928	0.943	0.931	0.948	0.94	0.956	0.938	0.946	1
181	PickleSong	2016-10-10	1	0.893	0.895	0.878	0.895	0.897	0.891	0.892	0.88	0.885
182	PickleSong	2016-10-12	0.893	1	0.966	0.955	0.969	0.97	0.96	0.968	0.959	0.953
183	PickleSong	2016-10-14	0.895	0.966	1	0.951	0.973	0.972	0.962	0.966	0.962	0.955
184	PickleSong	2016-10-17	0.878	0.955	0.951	1	0.954	0.955	0.947	0.951	0.942	0.937
185	PickleSong	2016-10-19	0.895	0.969	0.973	0.954	1	0.974	0.963	0.969	0.962	0.953
186	PickleSong	2016-10-21	0.897	0.97	0.972	0.955	0.974	1	0.965	0.971	0.961	0.954
187	PickleSong	2016-10-24	0.891	0.96	0.962	0.947	0.963	0.965	1	0.966	0.956	0.95
188	PickleSong	2016-10-26	0.892	0.968	0.966	0.951	0.969	0.971	0.966	1	0.966	0.959
189	PickleSong	2016-10-28	0.88	0.959	0.962	0.942	0.962	0.961	0.956	0.966	1	0.961
190	PickleSong	2016-10-31	0.885	0.953	0.955	0.937	0.953	0.954	0.95	0.959	0.961	1
191	Port Babbage	2016-10-10	1	0.806	0.792	0.787	0.776	0.777	0.774	0.768	0.774	0.771
192	Port Babbage	2016-10-12	0.806	1	0.964	0.948	0.939	0.94	0.926	0.922	0.932	0.922
193	Port Babbage	2016-10-14	0.792	0.964	1	0.965	0.959	0.959	0.936	0.933	0.943	0.93
194	Port Babbage	2016-10-17	0.787	0.948	0.965	1	0.957	0.961	0.931	0.923	0.934	0.922
195	Port Babbage	2016-10-19	0.776	0.939	0.959	0.957	1	0.978	0.954	0.96	0.955	0.934
196	Port Babbage	2016-10-21	0.777	0.94	0.959	0.961	0.978	1	0.949	0.953	0.96	0.939
197	Port Babbage	2016-10-24	0.774	0.926	0.936	0.931	0.954	0.949	1	0.968	0.971	0.946
198	Port Babbage	2016-10-26	0.768	0.922	0.933	0.923	0.96	0.953	0.968	1	0.977	0.952
199	Port Babbage	2016-10-28	0.774	0.932	0.943	0.934	0.955	0.96	0.971	0.977	1	0.97
200	Port Babbage	2016-10-31	0.771	0.922	0.93	0.922	0.934	0.939	0.946	0.952	0.97	1
201	PREFABRICA	2016-10-10	1.004	0.921	0.958	0.944	0.947	0.939	0.902	0.944	0.975	0.977
202	PREFABRICA	2016-10-12	0.921	1.002	0.946	0.964	0.964	0.964	0.965	0.963	0.916	0.918
203	PREFABRICA	2016-10-14	0.958	0.946	1.002	0.974	0.976	0.974	0.926	0.979	0.955	0.958
204	PREFABRICA	2016-10-17	0.944	0.964	0.974	1.002	0.988	0.99	0.947	0.99	0.936	0.944
205	PREFABRICA	2016-10-19	0.947	0.964	0.976	0.988	1.002	0.988	0.942	0.988	0.944	0.949
206	PREFABRICA	2016-10-21	0.939	0.964	0.974	0.99	0.988	1.002	0.946	0.993	0.942	0.949
207	PREFABRICA	2016-10-24	0.902	0.965	0.926	0.947	0.942	0.946	1.001	0.945	0.897	0.9

No	Region	ReferenceScan	2016-10-10	2016-10-12	2016-10-14	2016-10-17	2016-10-19	2016-10-21	2016-10-24	2016-10-26	2016-10-28	2016-10-31
208	PREFABRICA	2016-10-26	0.944	0.963	0.979	0.99	0.988	0.993	0.945	1.002	0.943	0.948
209	PREFABRICA	2016-10-28	0.975	0.916	0.955	0.936	0.944	0.942	0.897	0.943	1.002	0.986
210	PREFABRICA	2016-10-31	0.977	0.918	0.958	0.944	0.949	0.949	0.9	0.948	0.986	1.002
211	Quietly Tuesday	2016-10-10	1	0.967	0.959	0.964	0.968	0.955	0.965	0.966	0.962	0.958
212	Quietly Tuesday	2016-10-12	0.967	1	0.98	0.99	0.986	0.976	0.99	0.989	0.989	0.971
213	Quietly Tuesday	2016-10-14	0.959	0.98	1	0.984	0.983	0.987	0.981	0.978	0.979	0.98
214	Quietly Tuesday	2016-10-17	0.964	0.99	0.984	1	0.989	0.979	0.992	0.988	0.987	0.975
215	Quietly Tuesday	2016-10-19	0.968	0.986	0.983	0.989	1	0.977	0.987	0.987	0.986	0.978
216	Quietly Tuesday	2016-10-21	0.955	0.976	0.987	0.979	0.977	1	0.977	0.978	0.974	0.982
217	Quietly Tuesday	2016-10-24	0.965	0.99	0.981	0.992	0.987	0.977	1	0.988	0.988	0.975
218	Quietly Tuesday	2016-10-26	0.966	0.989	0.978	0.988	0.987	0.978	0.988	1	0.985	0.974
219	Quietly Tuesday	2016-10-28	0.962	0.989	0.979	0.987	0.986	0.974	0.988	0.985	1	0.97
220	Quietly Tuesday	2016-10-31	0.958	0.971	0.98	0.975	0.978	0.982	0.975	0.974	0.97	1
221	Trianwe	2016-10-10	1	0.958	0.96	0.965	0.949	0.957	0.958	0.953	0.964	0.961
222	Trianwe	2016-10-12	0.958	1	0.983	0.973	0.949	0.985	0.986	0.982	0.952	0.977
223	Trianwe	2016-10-14	0.96	0.983	1	0.977	0.956	0.988	0.989	0.985	0.956	0.981
224	Trianwe	2016-10-17	0.965	0.973	0.977	1	0.962	0.979	0.977	0.974	0.961	0.977
225	Trianwe	2016-10-19	0.949	0.949	0.956	0.962	1	0.955	0.955	0.949	0.953	0.953
226	Trianwe	2016-10-21	0.957	0.985	0.988	0.979	0.955	1	0.992	0.987	0.953	0.983
227	Trianwe	2016-10-24	0.958	0.986	0.989	0.977	0.955	0.992	1	0.986	0.952	0.982
228	Trianwe	2016-10-26	0.953	0.982	0.985	0.974	0.949	0.987	0.986	1	0.947	0.978
229	Trianwe	2016-10-28	0.964	0.952	0.956	0.961	0.953	0.953	0.952	0.947	1	0.952
230	Trianwe	2016-10-31	0.961	0.977	0.981	0.977	0.953	0.983	0.982	0.978	0.952	1
231	Triglav	2016-10-10	1.002	0.946	0.89	0.867	0.933	0.882	0.911	0.915	0.92	0.917
232	Triglav	2016-10-12	0.946	1.002	0.901	0.892	0.922	0.884	0.902	0.914	0.902	0.909
233	Triglav	2016-10-14	0.89	0.901	1.003	0.922	0.908	0.95	0.922	0.921	0.911	0.878
234	Triglav	2016-10-17	0.867	0.892	0.922	1.002	0.897	0.934	0.908	0.904	0.896	0.859
235	Triglav	2016-10-19	0.933	0.922	0.908	0.897	1.002	0.922	0.949	0.945	0.935	0.932
236	Triglav	2016-10-21	0.882	0.884	0.95	0.934	0.922	1.002	0.927	0.927	0.913	0.887
237	Triglav	2016-10-24	0.911	0.902	0.922	0.908	0.949	0.927	1.002	0.968	0.947	0.914
238	Triglav	2016-10-26	0.915	0.914	0.921	0.904	0.945	0.927	0.968	1.002	0.953	0.923
239	Triglav	2016-10-28	0.92	0.902	0.911	0.896	0.935	0.913	0.947	0.953	1.002	0.915
240	Triglav	2016-10-31	0.917	0.909	0.878	0.859	0.932	0.887	0.914	0.923	0.915	1.002
241	Yumix Prada	2016-10-10	1.006	0.672	0.673	0.682	0.678	0.67	0.538	0.683	0.691	0.672
242	Yumix Prada	2016-10-12	0.672	1.004	0.972	0.968	0.971	0.975	0.754	0.952	0.953	0.963
243	Yumix Prada	2016-10-14	0.673	0.972	1.005	0.982	0.984	0.987	0.762	0.964	0.97	0.981
244	Yumix Prada	2016-10-17	0.682	0.968	0.982	1.005	0.997	0.99	0.77	0.978	0.982	0.984
245	Yumix Prada	2016-10-19	0.678	0.971	0.984	0.997	1.005	0.991	0.771	0.98	0.981	0.985
246	Yumix Prada	2016-10-21	0.67	0.975	0.987	0.99	0.991	1.005	0.768	0.97	0.973	0.983
247	Yumix Prada	2016-10-24	0.538	0.754	0.762	0.77	0.771	0.768	1.004	0.757	0.76	0.763
248	Yumix Prada	2016-10-26	0.683	0.952	0.964	0.978	0.98	0.97	0.757	1.004	0.982	0.968
249	Yumix Prada	2016-10-28	0.691	0.953	0.97	0.982	0.981	0.973	0.76	0.982	1.005	0.972
250	Yumix Prada	2016-10-31	0.672	0.963	0.981	0.984	0.985	0.983	0.763	0.968	0.972	1.004

F.3 By-participant VE multinomial models, predicting manual classification patterns

ID	ModelType	Baseline	VECategory	TermInModel	Coefficient	StdErr	LRTest.ChiSq	DF	p	sig	McFadden	CoxSnell	Nagelkerke
1	With unclassifiable category	Unclassifiable	bored	Intercept	14406.2914	649.6607	NA	NA	NA	NA	0.9992	0.9921	1
1	With unclassifiable category	Unclassifiable	bored	Valence	10224.8379	1278.3908	10.3018	11	0.5035	ns.	0.9992	0.9921	1
1	With unclassifiable category	Unclassifiable	bored	Arousal	42663.8715	301.9751	22.0724	11	0.0238	*	0.9992	0.9921	1
1	With unclassifiable category	Unclassifiable	bored	Dominance	7606.7774	578.9182	-0.0961	11	1	ns.	0.9992	0.9921	1
1	With unclassifiable category	Unclassifiable	bored	Uncertainty	-35033.9056	116.0467	-0.0975	11	1	ns.	0.9992	0.9921	1
1	With unclassifiable category	Unclassifiable	bored	MBC:Cluster2	41175.5583	649.6607	-0.0279	11	1	ns.	0.9992	0.9921	1
1	With unclassifiable category	Unclassifiable	cute	Intercept	-23648.2553	NA	NA	NA	NA	NA	0.9992	0.9921	1
1	With unclassifiable category	Unclassifiable	cute	Valence	27754.8504	NA	10.3018	11	0.5035	ns.	0.9992	0.9921	1
1	With unclassifiable category	Unclassifiable	cute	Arousal	30807.9214	0	22.0724	11	0.0238	*	0.9992	0.9921	1
1	With unclassifiable category	Unclassifiable	cute	Dominance	75446.4062	0	-0.0961	11	1	ns.	0.9992	0.9921	1
1	With unclassifiable category	Unclassifiable	cute	Uncertainty	-16644.225	0	-0.0975	11	1	ns.	0.9992	0.9921	1
1	With unclassifiable category	Unclassifiable	cute	MBC:Cluster2	-21891.1096	0	-0.0279	11	1	ns.	0.9992	0.9921	1
1	With unclassifiable category	Unclassifiable	dream / imagination	Intercept	18511.8861	285.9105	NA	NA	NA	NA	0.9992	0.9921	1
1	With unclassifiable category	Unclassifiable	dream / imagination	Valence	10528.4454	436.0236	10.3018	11	0.5035	ns.	0.9992	0.9921	1
1	With unclassifiable category	Unclassifiable	dream / imagination	Arousal	42128.3522	13.298	22.0724	11	0.0238	*	0.9992	0.9921	1
1	With unclassifiable category	Unclassifiable	dream / imagination	Dominance	7723.582	511.9574	-0.0961	11	1	ns.	0.9992	0.9921	1
1	With unclassifiable category	Unclassifiable	dream / imagination	Uncertainty	-22684.448	59.7162	-0.0975	11	1	ns.	0.9992	0.9921	1
1	With unclassifiable category	Unclassifiable	dream / imagination	MBC:Cluster2	39217.6032	285.9105	-0.0279	11	1	ns.	0.9992	0.9921	1
1	With unclassifiable category	Unclassifiable	embarassed laugh & attraction	Intercept	9184.1324	1.9476	NA	NA	NA	NA	0.9992	0.9921	1
1	With unclassifiable category	Unclassifiable	embarassed laugh & attraction	Valence	-20416.1571	2.2112	10.3018	11	0.5035	ns.	0.9992	0.9921	1
1	With unclassifiable category	Unclassifiable	embarassed laugh & attraction	Arousal	392.1823	1.6298	22.0724	11	0.0238	*	0.9992	0.9921	1
1	With unclassifiable category	Unclassifiable	embarassed laugh & attraction	Dominance	-26886.5733	1.7994	-0.0961	11	1	ns.	0.9992	0.9921	1
1	With unclassifiable category	Unclassifiable	embarassed laugh & attraction	Uncertainty	-39453.8251	0.4463	-0.0975	11	1	ns.	0.9992	0.9921	1
1	With unclassifiable category	Unclassifiable	embarassed laugh & attraction	MBC:Cluster2	-8357.7336	0	-0.0279	11	1	ns.	0.9992	0.9921	1
1	With unclassifiable category	Unclassifiable	good memory	Intercept	-36618.6915	0	NA	NA	NA	NA	0.9992	0.9921	1
1	With unclassifiable category	Unclassifiable	good memory	Valence	83096.6198	0	10.3018	11	0.5035	ns.	0.9992	0.9921	1
1	With unclassifiable category	Unclassifiable	good memory	Arousal	-18642.1429	0	22.0724	11	0.0238	*	0.9992	0.9921	1
1	With unclassifiable category	Unclassifiable	good memory	Dominance	18417.1095	0	-0.0961	11	1	ns.	0.9992	0.9921	1
1	With unclassifiable category	Unclassifiable	good memory	Uncertainty	-12100.7705	0	-0.0975	11	1	ns.	0.9992	0.9921	1
1	With unclassifiable category	Unclassifiable	good memory	MBC:Cluster2	-39774.072	0	-0.0279	11	1	ns.	0.9992	0.9921	1
1	With unclassifiable category	Unclassifiable	impressive	Intercept	-37240.4738	0	NA	NA	NA	NA	0.9992	0.9921	1
1	With unclassifiable category	Unclassifiable	impressive	Valence	-19253.0434	0	10.3018	11	0.5035	ns.	0.9992	0.9921	1

ID	ModelType	Baseline	VECategory	TermInModel	Coefficient	StdErr	LRTest.ChiSq	DF	p	sig	McFadden	CoxSnell	Nagelkerke
1	With unclassifiable category	Unclassifiable	impressive	Arousal	-113087.4558	0	22.0724	11	0.0238	*	0.9992	0.9921	1
1	With unclassifiable category	Unclassifiable	impressive	Dominance	-67477.5935	0	-0.0961	11	1	ns.	0.9992	0.9921	1
1	With unclassifiable category	Unclassifiable	impressive	Uncertainty	-8092.6702	0	-0.0975	11	1	ns.	0.9992	0.9921	1
1	With unclassifiable category	Unclassifiable	impressive	MBC:Cluster2	6098.9827	0	-0.0279	11	1	ns.	0.9992	0.9921	1
1	With unclassifiable category	Unclassifiable	scary	Intercept	3637.7003	1.9476	NA	NA	NA	NA	0.9992	0.9921	1
1	With unclassifiable category	Unclassifiable	scary	Valence	-21522.4537	2.2112	10.3018	11	0.5035	ns.	0.9992	0.9921	1
1	With unclassifiable category	Unclassifiable	scary	Arousal	-17364.592	1.6298	22.0724	11	0.0238	*	0.9992	0.9921	1
1	With unclassifiable category	Unclassifiable	scary	Dominance	-49921.4479	1.7994	-0.0961	11	1	ns.	0.9992	0.9921	1
1	With unclassifiable category	Unclassifiable	scary	Uncertainty	-30171.2143	0.4463	-0.0975	11	1	ns.	0.9992	0.9921	1
1	With unclassifiable category	Unclassifiable	scary	MBC:Cluster2	7730.8801	0	-0.0279	11	1	ns.	0.9992	0.9921	1
1	With unclassifiable category	Unclassifiable	stressed	Intercept	37513.3765	0	NA	NA	NA	NA	0.9992	0.9921	1
1	With unclassifiable category	Unclassifiable	stressed	Valence	22563.74	0	10.3018	11	0.5035	ns.	0.9992	0.9921	1
1	With unclassifiable category	Unclassifiable	stressed	Arousal	47365.0529	0	22.0724	11	0.0238	*	0.9992	0.9921	1
1	With unclassifiable category	Unclassifiable	stressed	Dominance	36354.9918	0	-0.0961	11	1	ns.	0.9992	0.9921	1
1	With unclassifiable category	Unclassifiable	stressed	Uncertainty	-32913.761	0	-0.0975	11	1	ns.	0.9992	0.9921	1
1	With unclassifiable category	Unclassifiable	stressed	MBC:Cluster2	-29383.2741	0	-0.0279	11	1	ns.	0.9992	0.9921	1
1	With unclassifiable category	Unclassifiable	stressful	Intercept	10761.7776	364.7384	NA	NA	NA	NA	0.9992	0.9921	1
1	With unclassifiable category	Unclassifiable	stressful	Valence	10864.5038	843.293	10.3018	11	0.5035	ns.	0.9992	0.9921	1
1	With unclassifiable category	Unclassifiable	stressful	Arousal	43846.1373	288.8205	22.0724	11	0.0238	*	0.9992	0.9921	1
1	With unclassifiable category	Unclassifiable	stressful	Dominance	7807.563	71.1479	-0.0961	11	1	ns.	0.9992	0.9921	1
1	With unclassifiable category	Unclassifiable	stressful	Uncertainty	-25310.404	56.5684	-0.0975	11	1	ns.	0.9992	0.9921	1
1	With unclassifiable category	Unclassifiable	stressful	MBC:Cluster2	47166.7514	364.7384	-0.0279	11	1	ns.	0.9992	0.9921	1
1	With unclassifiable category	Unclassifiable	Worried	Intercept	16212.39	0	NA	NA	NA	NA	0.9992	0.9921	1
1	With unclassifiable category	Unclassifiable	Worried	Valence	-73625.4077	0	10.3018	11	0.5035	ns.	0.9992	0.9921	1
1	With unclassifiable category	Unclassifiable	Worried	Arousal	-38451.5802	0	22.0724	11	0.0238	*	0.9992	0.9921	1
1	With unclassifiable category	Unclassifiable	Worried	Dominance	23290.6329	0	-0.0961	11	1	ns.	0.9992	0.9921	1
1	With unclassifiable category	Unclassifiable	Worried	Uncertainty	-12466.7393	0	-0.0975	11	1	ns.	0.9992	0.9921	1
1	With unclassifiable category	Unclassifiable	Worried	MBC:Cluster2	-15023.1588	0	-0.0279	11	1	ns.	0.9992	0.9921	1
1	With unclassifiable category	Unclassifiable	zen	Intercept	1682.2373	0	NA	NA	NA	NA	0.9992	0.9921	1
1	With unclassifiable category	Unclassifiable	zen	Valence	-8299.1323	0	10.3018	11	0.5035	ns.	0.9992	0.9921	1
1	With unclassifiable category	Unclassifiable	zen	Arousal	-35712.9071	0	22.0724	11	0.0238	*	0.9992	0.9921	1
1	With unclassifiable category	Unclassifiable	zen	Dominance	-26175.2593	0	-0.0961	11	1	ns.	0.9992	0.9921	1
1	With unclassifiable category	Unclassifiable	zen	Uncertainty	-30923.348	0	-0.0975	11	1	ns.	0.9992	0.9921	1
1	With unclassifiable category	Unclassifiable	zen	MBC:Cluster2	18664.4109	0	-0.0279	11	1	ns.	0.9992	0.9921	1
2	With unclassifiable category	Unclassifiable	Alone	Intercept	5673.5945	28752.8349	NA	NA	NA	NA	1	0.988	1
2	With unclassifiable category	Unclassifiable	Alone	Valence	-5062.716	6376.2139	4.7969	9	0.8516	ns.	1	0.988	1
2	With unclassifiable category	Unclassifiable	Alone	Arousal	6376.7768	31293.5422	33.5557	9	0.0001	***	1	0.988	1
2	With unclassifiable category	Unclassifiable	Alone	Dominance	6517.4915	12165.8072	14.4362	9	0.1076	ns.	1	0.988	1
2	With unclassifiable category	Unclassifiable	Alone	Uncertainty	-12360.1055	6567.5581	-0.0001	9	1	ns.	1	0.988	1
2	With unclassifiable category	Unclassifiable	Alone	MBC:Cluster2	-5442.0272	NA	4.6844	9	0.8609	ns.	1	0.988	1
2	With unclassifiable category	Unclassifiable	excited	Intercept	2200.5869	0	NA	NA	NA	NA	1	0.988	1
2	With unclassifiable category	Unclassifiable	excited	Valence	-10196.9789	0	4.7969	9	0.8516	ns.	1	0.988	1
2	With unclassifiable category	Unclassifiable	excited	Arousal	1569.3366	0	33.5557	9	0.0001	***	1	0.988	1
2	With unclassifiable category	Unclassifiable	excited	Dominance	14376.6603	0	14.4362	9	0.1076	ns.	1	0.988	1
2	With unclassifiable category	Unclassifiable	excited	Uncertainty	-1699.6438	0	-0.0001	9	1	ns.	1	0.988	1
2	With unclassifiable category	Unclassifiable	excited	MBC:Cluster2	1117.9818	0	4.6844	9	0.8609	ns.	1	0.988	1
2	With unclassifiable category	Unclassifiable	happy	Intercept	-10728.2306	0	NA	NA	NA	NA	1	0.988	1
2	With unclassifiable category	Unclassifiable	happy	Valence	5954.6641	0	4.7969	9	0.8516	ns.	1	0.988	1
2	With unclassifiable category	Unclassifiable	happy	Arousal	32736.345	0	33.5557	9	0.0001	***	1	0.988	1
2	With unclassifiable category	Unclassifiable	happy	Dominance	31792.9496	0	14.4362	9	0.1076	ns.	1	0.988	1
2	With unclassifiable category	Unclassifiable	happy	Uncertainty	3360.3731	0	-0.0001	9	1	ns.	1	0.988	1

ID	ModelType	Baseline	VECategory	TermInModel	Coefficient	StdErr	LRTest.ChiSq	DF	p	sig	McFadden	CoxSnell	Nagelkerke
2	With unclassifiable category	Unclassifiable	happy	MBC:Cluster2	-4699.5753	0	4.6844	9	0.8609	ns.	1	0.988	1
2	With unclassifiable category	Unclassifiable	Hungry	Intercept	-671.0383	0	NA	NA	NA	NA	1	0.988	1
2	With unclassifiable category	Unclassifiable	Hungry	Valence	10088.4865	0	4.7969	9	0.8516	ns.	1	0.988	1
2	With unclassifiable category	Unclassifiable	Hungry	Arousal	14914.1571	0	33.5557	9	0.0001	***	1	0.988	1
2	With unclassifiable category	Unclassifiable	Hungry	Dominance	-4167.227	0	14.4362	9	0.1076	ns.	1	0.988	1
2	With unclassifiable category	Unclassifiable	Hungry	Uncertainty	-1065.9448	0	-0.0001	9	1	ns.	1	0.988	1
2	With unclassifiable category	Unclassifiable	Hungry	MBC:Cluster2	5838.7585	0	4.6844	9	0.8609	ns.	1	0.988	1
2	With unclassifiable category	Unclassifiable	Indifferent	Intercept	7541.6973	65.5066	NA	NA	NA	NA	1	0.988	1
2	With unclassifiable category	Unclassifiable	Indifferent	Valence	-9309.2852	92.3241	4.7969	9	0.8516	ns.	1	0.988	1
2	With unclassifiable category	Unclassifiable	Indifferent	Arousal	-1550.339	47.2573	33.5557	9	0.0001	***	1	0.988	1
2	With unclassifiable category	Unclassifiable	Indifferent	Dominance	12973.473	72.2828	14.4362	9	0.1076	ns.	1	0.988	1
2	With unclassifiable category	Unclassifiable	Indifferent	Uncertainty	5413.7351	15.0184	-0.0001	9	1	ns.	1	0.988	1
2	With unclassifiable category	Unclassifiable	Indifferent	MBC:Cluster2	-4324.4409	65.5066	4.6844	9	0.8609	ns.	1	0.988	1
2	With unclassifiable category	Unclassifiable	peaceful	Intercept	-12185.5956	0	NA	NA	NA	NA	1	0.988	1
2	With unclassifiable category	Unclassifiable	peaceful	Valence	-755.7854	0	4.7969	9	0.8516	ns.	1	0.988	1
2	With unclassifiable category	Unclassifiable	peaceful	Arousal	-31343.8395	0	33.5557	9	0.0001	***	1	0.988	1
2	With unclassifiable category	Unclassifiable	peaceful	Dominance	9091.9011	0	14.4362	9	0.1076	ns.	1	0.988	1
2	With unclassifiable category	Unclassifiable	peaceful	Uncertainty	2887.3803	0	-0.0001	9	1	ns.	1	0.988	1
2	With unclassifiable category	Unclassifiable	peaceful	MBC:Cluster2	-14876.1087	0	4.6844	9	0.8609	ns.	1	0.988	1
2	With unclassifiable category	Unclassifiable	relaxed	Intercept	5100.3982	65.5066	NA	NA	NA	NA	1	0.988	1
2	With unclassifiable category	Unclassifiable	relaxed	Valence	-8325.4911	92.3241	4.7969	9	0.8516	ns.	1	0.988	1
2	With unclassifiable category	Unclassifiable	relaxed	Arousal	14479.6226	47.2573	33.5557	9	0.0001	***	1	0.988	1
2	With unclassifiable category	Unclassifiable	relaxed	Dominance	14898.9151	72.2828	14.4362	9	0.1076	ns.	1	0.988	1
2	With unclassifiable category	Unclassifiable	relaxed	Uncertainty	12464.5097	15.0184	-0.0001	9	1	ns.	1	0.988	1
2	With unclassifiable category	Unclassifiable	relaxed	MBC:Cluster2	7799.6555	65.5066	4.6844	9	0.8609	ns.	1	0.988	1
2	With unclassifiable category	Unclassifiable	Uneasy	Intercept	2401.6212	28752.8349	NA	NA	NA	NA	1	0.988	1
2	With unclassifiable category	Unclassifiable	Uneasy	Valence	-4007.6099	6376.2139	4.7969	9	0.8516	ns.	1	0.988	1
2	With unclassifiable category	Unclassifiable	Uneasy	Arousal	11500.2629	31293.5422	33.5557	9	0.0001	***	1	0.988	1
2	With unclassifiable category	Unclassifiable	Uneasy	Dominance	4524.2928	12165.8072	14.4362	9	0.1076	ns.	1	0.988	1
2	With unclassifiable category	Unclassifiable	Uneasy	Uncertainty	-5069.7338	6567.5581	-0.0001	9	1	ns.	1	0.988	1
2	With unclassifiable category	Unclassifiable	Uneasy	MBC:Cluster2	-14146.3018	0	4.6844	9	0.8609	ns.	1	0.988	1
2	With unclassifiable category	Unclassifiable	WTF	Intercept	-19917.1218	0	NA	NA	NA	NA	1	0.988	1
2	With unclassifiable category	Unclassifiable	WTF	Valence	12311.901	0	4.7969	9	0.8516	ns.	1	0.988	1
2	With unclassifiable category	Unclassifiable	WTF	Arousal	30065.1563	0	33.5557	9	0.0001	***	1	0.988	1
2	With unclassifiable category	Unclassifiable	WTF	Dominance	-3824.4693	0	14.4362	9	0.1076	ns.	1	0.988	1
2	With unclassifiable category	Unclassifiable	WTF	Uncertainty	4433.7238	0	-0.0001	9	1	ns.	1	0.988	1
2	With unclassifiable category	Unclassifiable	WTF	MBC:Cluster2	3003.1023	0	4.6844	9	0.8609	ns.	1	0.988	1
3	With unclassifiable category	Unclassifiable	curious	Intercept	11.3846	52.0449	NA	NA	NA	NA	0.4847	0.8337	0.8548
3	With unclassifiable category	Unclassifiable	curious	Valence	-7.9115	24.8824	12.5458	6	0.0508	ns.	0.4847	0.8337	0.8548
3	With unclassifiable category	Unclassifiable	curious	Arousal	28.2209	68.0354	5.4092	6	0.4925	ns.	0.4847	0.8337	0.8548
3	With unclassifiable category	Unclassifiable	curious	Dominance	27.8478	64.9365	12.9333	6	0.0441	*	0.4847	0.8337	0.8548
3	With unclassifiable category	Unclassifiable	curious	Uncertainty	88.4496	75.8677	17.531	6	0.0075	**	0.4847	0.8337	0.8548
3	With unclassifiable category	Unclassifiable	curious	MBC:Cluster2	56.2305	87.3482	9.5075	6	0.147	ns.	0.4847	0.8337	0.8548
3	With unclassifiable category	Unclassifiable	empty	Intercept	-24.6207	58.1002	NA	NA	NA	NA	0.4847	0.8337	0.8548
3	With unclassifiable category	Unclassifiable	empty	Valence	-8.4253	24.8669	12.5458	6	0.0508	ns.	0.4847	0.8337	0.8548
3	With unclassifiable category	Unclassifiable	empty	Arousal	26.7003	68.0165	5.4092	6	0.4925	ns.	0.4847	0.8337	0.8548
3	With unclassifiable category	Unclassifiable	empty	Dominance	26.1478	64.9213	12.9333	6	0.0441	*	0.4847	0.8337	0.8548
3	With unclassifiable category	Unclassifiable	empty	Uncertainty	-66.9162	134.7849	17.531	6	0.0075	**	0.4847	0.8337	0.8548
3	With unclassifiable category	Unclassifiable	empty	MBC:Cluster2	58.3731	87.3738	9.5075	6	0.147	ns.	0.4847	0.8337	0.8548
3	With unclassifiable category	Unclassifiable	end	Intercept	1.719	184.831	NA	NA	NA	NA	0.4847	0.8337	0.8548
3	With unclassifiable category	Unclassifiable	end	Valence	-11.4383	47.7389	12.5458	6	0.0508	ns.	0.4847	0.8337	0.8548
3	With unclassifiable category	Unclassifiable	end	Arousal	-12.5794	199.3251	5.4092	6	0.4925	ns.	0.4847	0.8337	0.8548

ID	ModelType	Baseline	VECategory	TermInModel	Coefficient	StdErr	LRTest.ChiSq	DF	p	sig	McFadden	CoxSnell	Nagelkerke
3	With unclassifiable category	Unclassifiable	end	Dominance	27.5741	102.4092	12.9333	6	0.0441	*	0.4847	0.8337	0.8548
3	With unclassifiable category	Unclassifiable	end	Uncertainty	-100.6341	41.3412	17.531	6	0.0075	**	0.4847	0.8337	0.8548
3	With unclassifiable category	Unclassifiable	end	MBC:Cluster2	-32.9242	434.6297	9.5075	6	0.147	ns.	0.4847	0.8337	0.8548
3	With unclassifiable category	Unclassifiable	observer	Intercept	8.9452	55.2226	NA	NA	NA	NA	0.4847	0.8337	0.8548
3	With unclassifiable category	Unclassifiable	observer	Valence	48.3533	61.8454	12.5458	6	0.0508	ns.	0.4847	0.8337	0.8548
3	With unclassifiable category	Unclassifiable	observer	Arousal	49.2608	74.5075	5.4092	6	0.4925	ns.	0.4847	0.8337	0.8548
3	With unclassifiable category	Unclassifiable	observer	Dominance	-16.7168	79.255	12.9333	6	0.0441	*	0.4847	0.8337	0.8548
3	With unclassifiable category	Unclassifiable	observer	Uncertainty	255.2459	174.5176	17.531	6	0.0075	**	0.4847	0.8337	0.8548
3	With unclassifiable category	Unclassifiable	observer	MBC:Cluster2	87.2496	96.447	9.5075	6	0.147	ns.	0.4847	0.8337	0.8548
3	With unclassifiable category	Unclassifiable	pleasant	Intercept	-39.7437	57.0014	NA	NA	NA	NA	0.4847	0.8337	0.8548
3	With unclassifiable category	Unclassifiable	pleasant	Valence	-6.4263	25.0755	12.5458	6	0.0508	ns.	0.4847	0.8337	0.8548
3	With unclassifiable category	Unclassifiable	pleasant	Arousal	28.2515	68.0463	5.4092	6	0.4925	ns.	0.4847	0.8337	0.8548
3	With unclassifiable category	Unclassifiable	pleasant	Dominance	24.1945	64.9034	12.9333	6	0.0441	*	0.4847	0.8337	0.8548
3	With unclassifiable category	Unclassifiable	pleasant	Uncertainty	-120.9605	124.4822	17.531	6	0.0075	**	0.4847	0.8337	0.8548
3	With unclassifiable category	Unclassifiable	pleasant	MBC:Cluster2	63.1407	87.4168	9.5075	6	0.147	ns.	0.4847	0.8337	0.8548
3	With unclassifiable category	Unclassifiable	wow	Intercept	-9.0525	53.8919	NA	NA	NA	NA	0.4847	0.8337	0.8548
3	With unclassifiable category	Unclassifiable	wow	Valence	-1.7585	25.8104	12.5458	6	0.0508	ns.	0.4847	0.8337	0.8548
3	With unclassifiable category	Unclassifiable	wow	Arousal	24.4104	68.1146	5.4092	6	0.4925	ns.	0.4847	0.8337	0.8548
3	With unclassifiable category	Unclassifiable	wow	Dominance	19.9615	64.7585	12.9333	6	0.0441	*	0.4847	0.8337	0.8548
3	With unclassifiable category	Unclassifiable	wow	Uncertainty	-13.9691	85.3289	17.531	6	0.0075	**	0.4847	0.8337	0.8548
3	With unclassifiable category	Unclassifiable	wow	MBC:Cluster2	52.2212	87.5702	9.5075	6	0.147	ns.	0.4847	0.8337	0.8548
4	Without unclassifiable category	enchanted	Creepy	Intercept	1385.1716	57477.398	NA	NA	NA	NA	1	0.9716	1
4	Without unclassifiable category	enchanted	Creepy	Valence	-456.3995	144116.5273	12.0255	6	0.0614	ns.	1	0.9716	1
4	Without unclassifiable category	enchanted	Creepy	Arousal	3058.137	93200.0342	19.1941	6	0.0038	**	1	0.9716	1
4	Without unclassifiable category	enchanted	Creepy	Dominance	1115.8928	154843.509	8.5805	6	0.1986	ns.	1	0.9716	1
4	Without unclassifiable category	enchanted	Creepy	Uncertainty	-665.2389	15913.6308	24.8135	6	0.0004	***	1	0.9716	1
4	Without unclassifiable category	enchanted	Creepy	MBC:Cluster2	-78.7471	38049.6197	17.5397	6	0.0075	**	1	0.9716	1
4	Without unclassifiable category	enchanted	sex	Intercept	-1397.6409	34154.6015	NA	NA	NA	NA	1	0.9716	1
4	Without unclassifiable category	enchanted	sex	Valence	-1075.3052	58444.1005	12.0255	6	0.0614	ns.	1	0.9716	1
4	Without unclassifiable category	enchanted	sex	Arousal	2374.2362	43886.0836	19.1941	6	0.0038	**	1	0.9716	1
4	Without unclassifiable category	enchanted	sex	Dominance	-1613.8677	25623.1213	8.5805	6	0.1986	ns.	1	0.9716	1
4	Without unclassifiable category	enchanted	sex	Uncertainty	-1131.4284	7833.8985	24.8135	6	0.0004	***	1	0.9716	1
4	Without unclassifiable category	enchanted	sex	MBC:Cluster2	-1573.765	0	17.5397	6	0.0075	**	1	0.9716	1
4	Without unclassifiable category	enchanted	City Life	Intercept	4730.364	4187.4328	NA	NA	NA	NA	1	0.9716	1
4	Without unclassifiable category	enchanted	City Life	Valence	69.1076	58890.4405	12.0255	6	0.0614	ns.	1	0.9716	1
4	Without unclassifiable category	enchanted	City Life	Arousal	860.4157	6190.5087	19.1941	6	0.0038	**	1	0.9716	1
4	Without unclassifiable category	enchanted	City Life	Dominance	707.7182	49169.6344	8.5805	6	0.1986	ns.	1	0.9716	1
4	Without unclassifiable category	enchanted	City Life	Uncertainty	5029.0373	69885.7732	24.8135	6	0.0004	***	1	0.9716	1
4	Without unclassifiable category	enchanted	City Life	MBC:Cluster2	-3695.5971	10615.9341	17.5397	6	0.0075	**	1	0.9716	1
4	Without unclassifiable category	enchanted	Depressing	Intercept	1017.8767	4797.1472	NA	NA	NA	NA	1	0.9716	1
4	Without unclassifiable category	enchanted	Depressing	Valence	-2788.6633	4623.5604	12.0255	6	0.0614	ns.	1	0.9716	1
4	Without unclassifiable category	enchanted	Depressing	Arousal	-2884.7523	692.1438	19.1941	6	0.0038	**	1	0.9716	1
4	Without unclassifiable category	enchanted	Depressing	Dominance	-1730.7361	3621.7822	8.5805	6	0.1986	ns.	1	0.9716	1
4	Without unclassifiable category	enchanted	Depressing	Uncertainty	2611.0345	652.7803	24.8135	6	0.0004	***	1	0.9716	1
4	Without unclassifiable category	enchanted	Depressing	MBC:Cluster2	-3438.2726	0	17.5397	6	0.0075	**	1	0.9716	1
4	Without unclassifiable category	enchanted	Fun	Intercept	-2829.9577	20595.288	NA	NA	NA	NA	1	0.9716	1
4	Without unclassifiable category	enchanted	Fun	Valence	-350.3219	11773.5373	12.0255	6	0.0614	ns.	1	0.9716	1
4	Without unclassifiable category	enchanted	Fun	Arousal	4004.1149	65093.8052	19.1941	6	0.0038	**	1	0.9716	1
4	Without unclassifiable category	enchanted	Fun	Dominance	893.5013	16522.3023	8.5805	6	0.1986	ns.	1	0.9716	1
4	Without unclassifiable category	enchanted	Fun	Uncertainty	-1054.5502	80532.6651	24.8135	6	0.0004	***	1	0.9716	1
4	Without unclassifiable category	enchanted	Fun	MBC:Cluster2	4839.724	20595.288	17.5397	6	0.0075	**	1	0.9716	1
4	Without unclassifiable category	enchanted	terrified	Intercept	-3331.7681	0	NA	NA	NA	NA	1	0.9716	1

ID	ModelType	Baseline	VECategory	TermInModel	Coefficient	StdErr	LRTest.ChiSq	DF	p	sig	McFadden	CoxSnell	Nagelkerke
4	Without unclassifiable category	enchanted	terrified	Valence	-864.7075	0	12.0255	6	0.0614	ns.	1	0.9716	1
4	Without unclassifiable category	enchanted	terrified	Arousal	3367.2677	0	19.1941	6	0.0038	**	1	0.9716	1
4	Without unclassifiable category	enchanted	terrified	Dominance	-1847.3611	0	8.5805	6	0.1986	ns.	1	0.9716	1
4	Without unclassifiable category	enchanted	terrified	Uncertainty	1191.4695	0	24.8135	6	0.0004	***	1	0.9716	1
4	Without unclassifiable category	enchanted	terrified	MBC:Cluster2	1248.3518	0	17.5397	6	0.0075	**	1	0.9716	1
5	Without unclassifiable category	interesting / curious	bored	Intercept	-76.9282	258.9064	NA	NA	NA	NA	0.7384	0.9405	0.9616
5	Without unclassifiable category	interesting / curious	bored	Valence	-4.037	12.0918	18.0501	6	0.0061	**	0.7384	0.9405	0.9616
5	Without unclassifiable category	interesting / curious	bored	Arousal	1.4037	8.3181	8.7337	6	0.1891	ns.	0.7384	0.9405	0.9616
5	Without unclassifiable category	interesting / curious	bored	Dominance	-3.2835	5.4461	11.4001	6	0.0768	ns.	0.7384	0.9405	0.9616
5	Without unclassifiable category	interesting / curious	bored	Uncertainty	15.2482	55.6601	1.708	6	0.9445	ns.	0.7384	0.9405	0.9616
5	Without unclassifiable category	interesting / curious	bored	MBC:Cluster2	86.3392	268.8023	5.5494	6	0.4755	ns.	0.7384	0.9405	0.9616
5	Without unclassifiable category	interesting / curious	cool / interesting	Intercept	-4.8061	6.9738	NA	NA	NA	NA	0.7384	0.9405	0.9616
5	Without unclassifiable category	interesting / curious	cool / interesting	Valence	-6.3896	6.9001	18.0501	6	0.0061	**	0.7384	0.9405	0.9616
5	Without unclassifiable category	interesting / curious	cool / interesting	Arousal	-2.3444	3.6022	8.7337	6	0.1891	ns.	0.7384	0.9405	0.9616
5	Without unclassifiable category	interesting / curious	cool / interesting	Dominance	1.1313	4.1752	11.4001	6	0.0768	ns.	0.7384	0.9405	0.9616
5	Without unclassifiable category	interesting / curious	cool / interesting	Uncertainty	-0.0119	1.7062	1.708	6	0.9445	ns.	0.7384	0.9405	0.9616
5	Without unclassifiable category	interesting / curious	cool / interesting	MBC:Cluster2	7.2455	12.2297	5.5494	6	0.4755	ns.	0.7384	0.9405	0.9616
5	Without unclassifiable category	interesting / curious	creepy	Intercept	-163.581	5.9033	NA	NA	NA	NA	0.7384	0.9405	0.9616
5	Without unclassifiable category	interesting / curious	creepy	Valence	-86.249	122.461	18.0501	6	0.0061	**	0.7384	0.9405	0.9616
5	Without unclassifiable category	interesting / curious	creepy	Arousal	121.6933	695.5371	8.7337	6	0.1891	ns.	0.7384	0.9405	0.9616
5	Without unclassifiable category	interesting / curious	creepy	Dominance	82.6769	983.9554	11.4001	6	0.0768	ns.	0.7384	0.9405	0.9616
5	Without unclassifiable category	interesting / curious	creepy	Uncertainty	5.5798	1.4281	1.708	6	0.9445	ns.	0.7384	0.9405	0.9616
5	Without unclassifiable category	interesting / curious	creepy	MBC:Cluster2	26.1774	0	5.5494	6	0.4755	ns.	0.7384	0.9405	0.9616
5	Without unclassifiable category	interesting / curious	cute / happy	Intercept	-58.4928	400.3727	NA	NA	NA	NA	0.7384	0.9405	0.9616
5	Without unclassifiable category	interesting / curious	cute / happy	Valence	11.316	16.8616	18.0501	6	0.0061	**	0.7384	0.9405	0.9616
5	Without unclassifiable category	interesting / curious	cute / happy	Arousal	40.3031	70.7542	8.7337	6	0.1891	ns.	0.7384	0.9405	0.9616
5	Without unclassifiable category	interesting / curious	cute / happy	Dominance	3.0384	6.971	11.4001	6	0.0768	ns.	0.7384	0.9405	0.9616
5	Without unclassifiable category	interesting / curious	cute / happy	Uncertainty	-5.8731	1394.4801	1.708	6	0.9445	ns.	0.7384	0.9405	0.9616

ID	ModelType	Baseline	VECategory	TermInModel	Coefficient	StdErr	LRTest.ChiSq	DF	p	sig	McFadden	CoxSnell	Nagelkerke
5	Without unclassifiable category	interesting / curious	cute / happy	MBC:Cluster2	72.5518	155.6517	5.5494	6	0.4755	ns.	0.7384	0.9405	0.9616
5	Without unclassifiable category	interesting / curious	pretty / relaxing	Intercept	-8.3721	200.5967	NA	NA	NA	NA	0.7384	0.9405	0.9616
5	Without unclassifiable category	interesting / curious	pretty / relaxing	Valence	40.6554	41.144	18.0501	6	0.0061	**	0.7384	0.9405	0.9616
5	Without unclassifiable category	interesting / curious	pretty / relaxing	Arousal	2.2647	8.0626	8.7337	6	0.1891	ns.	0.7384	0.9405	0.9616
5	Without unclassifiable category	interesting / curious	pretty / relaxing	Dominance	-18.4597	16.9843	11.4001	6	0.0768	ns.	0.7384	0.9405	0.9616
5	Without unclassifiable category	interesting / curious	pretty / relaxing	Uncertainty	-36.245	854.9972	1.708	6	0.9445	ns.	0.7384	0.9405	0.9616
5	Without unclassifiable category	interesting / curious	pretty / relaxing	MBC:Cluster2	-17.8842	34.3939	5.5494	6	0.4755	ns.	0.7384	0.9405	0.9616
5	Without unclassifiable category	interesting / curious	uncomfortable / creepy	Intercept	-121.8936	85.7631	NA	NA	NA	NA	0.7384	0.9405	0.9616
5	Without unclassifiable category	interesting / curious	uncomfortable / creepy	Valence	-20.0442	17.2094	18.0501	6	0.0061	**	0.7384	0.9405	0.9616
5	Without unclassifiable category	interesting / curious	uncomfortable / creepy	Arousal	-35.9063	32.6406	8.7337	6	0.1891	ns.	0.7384	0.9405	0.9616
5	Without unclassifiable category	interesting / curious	uncomfortable / creepy	Dominance	-116.3054	104.7977	11.4001	6	0.0768	ns.	0.7384	0.9405	0.9616
5	Without unclassifiable category	interesting / curious	uncomfortable / creepy	Uncertainty	-92.4287	19.2065	1.708	6	0.9445	ns.	0.7384	0.9405	0.9616
5	Without unclassifiable category	interesting / curious	uncomfortable / creepy	MBC:Cluster2	-17.2473	0	5.5494	6	0.4755	ns.	0.7384	0.9405	0.9616
6	With unclassifiable category	Unclassifiable	relaxed	Intercept	2.6112	117.2002	NA	NA	NA	NA	0.5301	0.7944	0.8367
6	With unclassifiable category	Unclassifiable	relaxed	Valence	-95.0932	32.4267	8.9831	4	0.0615	ns.	0.5301	0.7944	0.8367
6	With unclassifiable category	Unclassifiable	relaxed	Arousal	-0.833	31.7936	3.9635	4	0.411	ns.	0.5301	0.7944	0.8367
6	With unclassifiable category	Unclassifiable	relaxed	Dominance	45.327	37.0879	4.1984	4	0.3798	ns.	0.5301	0.7944	0.8367
6	With unclassifiable category	Unclassifiable	relaxed	Uncertainty	17.491	48.27	3.457	4	0.4844	ns.	0.5301	0.7944	0.8367
6	With unclassifiable category	Unclassifiable	relaxed	MBC:Cluster2	68.2356	118.9269	4.4276	4	0.3512	ns.	0.5301	0.7944	0.8367
6	With unclassifiable category	Unclassifiable	confused	Intercept	50.3547	45.1798	NA	NA	NA	NA	0.5301	0.7944	0.8367
6	With unclassifiable category	Unclassifiable	confused	Valence	-90.1796	32.404	8.9831	4	0.0615	ns.	0.5301	0.7944	0.8367
6	With unclassifiable category	Unclassifiable	confused	Arousal	5.3473	31.8109	3.9635	4	0.411	ns.	0.5301	0.7944	0.8367
6	With unclassifiable category	Unclassifiable	confused	Dominance	39.9571	37.0864	4.1984	4	0.3798	ns.	0.5301	0.7944	0.8367
6	With unclassifiable category	Unclassifiable	confused	Uncertainty	-27.349	49.2466	3.457	4	0.4844	ns.	0.5301	0.7944	0.8367
6	With unclassifiable category	Unclassifiable	confused	MBC:Cluster2	14.2778	46.1249	4.4276	4	0.3512	ns.	0.5301	0.7944	0.8367
6	With unclassifiable category	Unclassifiable	Creepy	Intercept	43.6049	45.2889	NA	NA	NA	NA	0.5301	0.7944	0.8367
6	With unclassifiable category	Unclassifiable	Creepy	Valence	-93.3932	32.4065	8.9831	4	0.0615	ns.	0.5301	0.7944	0.8367
6	With unclassifiable category	Unclassifiable	Creepy	Arousal	8.2819	31.846	3.9635	4	0.411	ns.	0.5301	0.7944	0.8367
6	With unclassifiable category	Unclassifiable	Creepy	Dominance	40.3815	37.0895	4.1984	4	0.3798	ns.	0.5301	0.7944	0.8367
6	With unclassifiable category	Unclassifiable	Creepy	Uncertainty	-29.455	48.3272	3.457	4	0.4844	ns.	0.5301	0.7944	0.8367
6	With unclassifiable category	Unclassifiable	Creepy	MBC:Cluster2	24.0317	45.7769	4.4276	4	0.3512	ns.	0.5301	0.7944	0.8367
6	With unclassifiable category	Unclassifiable	Slightly Intimidating	Intercept	41.6478	46.1391	NA	NA	NA	NA	0.5301	0.7944	0.8367
6	With unclassifiable category	Unclassifiable	Slightly Intimidating	Valence	-90.3797	32.7163	8.9831	4	0.0615	ns.	0.5301	0.7944	0.8367
6	With unclassifiable category	Unclassifiable	Slightly Intimidating	Arousal	2.0158	31.7875	3.9635	4	0.411	ns.	0.5301	0.7944	0.8367
6	With unclassifiable category	Unclassifiable	Slightly Intimidating	Dominance	41.0249	37.1341	4.1984	4	0.3798	ns.	0.5301	0.7944	0.8367

ID	ModelType	Baseline	VECategory	TermInModel	Coefficient	StdErr	LRTest.ChiSq	DF	p	sig	McFadden	CoxSnell	Nagelkerke
6	With unclassifiable category	Unclassifiable	Slightly Intimidating	Uncertainty	10.522	52.8243	3.457	4	0.4844	ns.	0.5301	0.7944	0.8367
6	With unclassifiable category	Unclassifiable	Slightly Intimidating	MBC:Cluster2	29.4179	50.7516	4.4276	4	0.3512	ns.	0.5301	0.7944	0.8367
7	Without unclassifiable category	tourist attractions	peaceful	Intercept	-3967.6746	0	NA	NA	NA	NA	1	0.9819	1
7	Without unclassifiable category	tourist attractions	peaceful	Valence	481.468	0	0.0125	7	1	ns.	1	0.9819	1
7	Without unclassifiable category	tourist attractions	peaceful	Arousal	4352.441	0	43.9692	7	0	***	1	0.9819	1
7	Without unclassifiable category	tourist attractions	peaceful	Dominance	8551.6359	0	36.3333	7	0	***	1	0.9819	1
7	Without unclassifiable category	tourist attractions	peaceful	Uncertainty	871.6421	0	0	7	1	ns.	1	0.9819	1
7	Without unclassifiable category	tourist attractions	peaceful	MBC:Cluster2	-4393.4716	0	0	7	1	ns.	1	0.9819	1
7	Without unclassifiable category	tourist attractions	brothels	Intercept	-845.0559	1446.6347	NA	NA	NA	NA	1	0.9819	1
7	Without unclassifiable category	tourist attractions	brothels	Valence	-1635.3966	2097.4636	0.0125	7	1	ns.	1	0.9819	1
7	Without unclassifiable category	tourist attractions	brothels	Arousal	1371.6059	1310.786	43.9692	7	0	***	1	0.9819	1
7	Without unclassifiable category	tourist attractions	brothels	Dominance	132.1879	1336.5961	36.3333	7	0	***	1	0.9819	1
7	Without unclassifiable category	tourist attractions	brothels	Uncertainty	-2572.7303	331.6459	0	7	1	ns.	1	0.9819	1
7	Without unclassifiable category	tourist attractions	brothels	MBC:Cluster2	-4595.7352	0	0	7	1	ns.	1	0.9819	1
7	Without unclassifiable category	tourist attractions	events	Intercept	-687.3242	764.6332	NA	NA	NA	NA	1	0.9819	1
7	Without unclassifiable category	tourist attractions	events	Valence	76.8018	2394.7001	0.0125	7	1	ns.	1	0.9819	1
7	Without unclassifiable category	tourist attractions	events	Arousal	2829.9746	1497.9056	43.9692	7	0	***	1	0.9819	1
7	Without unclassifiable category	tourist attractions	events	Dominance	424.066	2479.5145	36.3333	7	0	***	1	0.9819	1
7	Without unclassifiable category	tourist attractions	events	Uncertainty	-1163.3119	139.2838	0	7	1	ns.	1	0.9819	1
7	Without unclassifiable category	tourist attractions	events	MBC:Cluster2	1872.1987	764.6332	0	7	1	ns.	1	0.9819	1
7	Without unclassifiable category	tourist attractions	Garden	Intercept	-2956.6385	344.5598	NA	NA	NA	NA	1	0.9819	1
7	Without unclassifiable category	tourist attractions	Garden	Valence	-1463.5872	7827.7739	0.0125	7	1	ns.	1	0.9819	1
7	Without unclassifiable category	tourist attractions	Garden	Arousal	-9356.4122	1196.3267	43.9692	7	0	***	1	0.9819	1
7	Without unclassifiable category	tourist attractions	Garden	Dominance	1850.4679	3302.6676	36.3333	7	0	***	1	0.9819	1
7	Without unclassifiable category	tourist attractions	Garden	Uncertainty	-354.9729	258.1995	0	7	1	ns.	1	0.9819	1
7	Without unclassifiable category	tourist attractions	Garden	MBC:Cluster2	-3871.731	344.5598	0	7	1	ns.	1	0.9819	1

ID	ModelType	Baseline	VECategory	TermInModel	Coefficient	StdErr	LRTest.ChiSq	DF	p	sig	McFadden	CoxSnell	Nagelkerke
7	Without unclassifiable category	tourist attractions	gothic	Intercept	-6444.7765	0	NA	NA	NA	NA	1	0.9819	1
7	Without unclassifiable category	tourist attractions	gothic	Valence	24.3166	0	0.0125	7	1	ns.	1	0.9819	1
7	Without unclassifiable category	tourist attractions	gothic	Arousal	-7586.6208	0	43.9692	7	0	***	1	0.9819	1
7	Without unclassifiable category	tourist attractions	gothic	Dominance	-14655.8718	0	36.3333	7	0	***	1	0.9819	1
7	Without unclassifiable category	tourist attractions	gothic	Uncertainty	722.3989	0	0	7	1	ns.	1	0.9819	1
7	Without unclassifiable category	tourist attractions	gothic	MBC:Cluster2	-318.1498	0	0	7	1	ns.	1	0.9819	1
7	Without unclassifiable category	tourist attractions	strange	Intercept	-2611.4259	0	NA	NA	NA	NA	1	0.9819	1
7	Without unclassifiable category	tourist attractions	strange	Valence	5733.5588	0	0.0125	7	1	ns.	1	0.9819	1
7	Without unclassifiable category	tourist attractions	strange	Arousal	13650.7152	0	43.9692	7	0	***	1	0.9819	1
7	Without unclassifiable category	tourist attractions	strange	Dominance	6547.0744	0	36.3333	7	0	***	1	0.9819	1
7	Without unclassifiable category	tourist attractions	strange	Uncertainty	541.043	0	0	7	1	ns.	1	0.9819	1
7	Without unclassifiable category	tourist attractions	strange	MBC:Cluster2	-4447.3246	0	0	7	1	ns.	1	0.9819	1
7	Without unclassifiable category	tourist attractions	zombie towns	Intercept	1764.7351	1446.6347	NA	NA	NA	NA	1	0.9819	1
7	Without unclassifiable category	tourist attractions	zombie towns	Valence	-1329.5264	2097.4636	0.0125	7	1	ns.	1	0.9819	1
7	Without unclassifiable category	tourist attractions	zombie towns	Arousal	1391.9311	1310.786	43.9692	7	0	***	1	0.9819	1
7	Without unclassifiable category	tourist attractions	zombie towns	Dominance	1955.7187	1336.5961	36.3333	7	0	***	1	0.9819	1
7	Without unclassifiable category	tourist attractions	zombie towns	Uncertainty	-308.8797	331.6459	0	7	1	ns.	1	0.9819	1
7	Without unclassifiable category	tourist attractions	zombie towns	MBC:Cluster2	-2972.074	0	0	7	1	ns.	1	0.9819	1
8	With unclassifiable category	Unclassifiable	Beachside	Intercept	-5827.5015	0	NA	NA	NA	NA	1	0.9905	1
8	With unclassifiable category	Unclassifiable	Beachside	Valence	-11724.2494	NA	0	11	1	ns.	1	0.9905	1
8	With unclassifiable category	Unclassifiable	Beachside	Arousal	-1109.4964	0	0	11	1	ns.	1	0.9905	1
8	With unclassifiable category	Unclassifiable	Beachside	Dominance	4756.7941	NA	21.5599	11	0.028	*	1	0.9905	1
8	With unclassifiable category	Unclassifiable	Beachside	Uncertainty	841.5946	0	0	11	1	ns.	1	0.9905	1
8	With unclassifiable category	Unclassifiable	Beachside	MBC:Cluster2	-3783.05	NA	0	11	1	ns.	1	0.9905	1
8	With unclassifiable category	Unclassifiable	Castle	Intercept	-176.6166	3903.8342	NA	NA	NA	NA	1	0.9905	1
8	With unclassifiable category	Unclassifiable	Castle	Valence	-1384.547	21715.9927	0	11	1	ns.	1	0.9905	1
8	With unclassifiable category	Unclassifiable	Castle	Arousal	-1724.2431	4340.5556	0	11	1	ns.	1	0.9905	1
8	With unclassifiable category	Unclassifiable	Castle	Dominance	-3957.5236	24314.8782	21.5599	11	0.028	*	1	0.9905	1
8	With unclassifiable category	Unclassifiable	Castle	Uncertainty	228.8577	3516.3622	0	11	1	ns.	1	0.9905	1
8	With unclassifiable category	Unclassifiable	Castle	MBC:Cluster2	3839.6876	4513.3771	0	11	1	ns.	1	0.9905	1
8	With unclassifiable category	Unclassifiable	Cathedral	Intercept	-3409.3429	0	NA	NA	NA	NA	1	0.9905	1
8	With unclassifiable category	Unclassifiable	Cathedral	Valence	1323.0752	0	0	11	1	ns.	1	0.9905	1
8	With unclassifiable category	Unclassifiable	Cathedral	Arousal	-11801.5914	0	0	11	1	ns.	1	0.9905	1
8	With unclassifiable category	Unclassifiable	Cathedral	Dominance	-17271.6485	0	21.5599	11	0.028	*	1	0.9905	1

ID	ModelType	Baseline	VECategory	TermInModel	Coefficient	StdErr	LRTest.ChiSq	DF	p	sig	McFadden	CoxSnell	Nagelkerke
8	With unclassifiable category	Unclassifiable	Cathedral	Uncertainty	75.9453	0	0	11	1	ns.	1	0.9905	1
8	With unclassifiable category	Unclassifiable	Cathedral	MBC:Cluster2	719.3806	0	0	11	1	ns.	1	0.9905	1
8	With unclassifiable category	Unclassifiable	City	Intercept	260.1971	NA	NA	NA	NA	NA	1	0.9905	1
8	With unclassifiable category	Unclassifiable	City	Valence	-7500.6982	NA	0	11	1	ns.	1	0.9905	1
8	With unclassifiable category	Unclassifiable	City	Arousal	-16702.8033	NA	0	11	1	ns.	1	0.9905	1
8	With unclassifiable category	Unclassifiable	City	Dominance	-9288.3738	NA	21.5599	11	0.028	*	1	0.9905	1
8	With unclassifiable category	Unclassifiable	City	Uncertainty	3077.9302	NA	0	11	1	ns.	1	0.9905	1
8	With unclassifiable category	Unclassifiable	City	MBC:Cluster2	-2062.9881	NA	0	11	1	ns.	1	0.9905	1
8	With unclassifiable category	Unclassifiable	Dinner	Intercept	-1792.5199	0	NA	NA	NA	NA	1	0.9905	1
8	With unclassifiable category	Unclassifiable	Dinner	Valence	3743.1538	0	0	11	1	ns.	1	0.9905	1
8	With unclassifiable category	Unclassifiable	Dinner	Arousal	-1509.9345	0	0	11	1	ns.	1	0.9905	1
8	With unclassifiable category	Unclassifiable	Dinner	Dominance	-7688.7653	0	21.5599	11	0.028	*	1	0.9905	1
8	With unclassifiable category	Unclassifiable	Dinner	Uncertainty	662.8605	0	0	11	1	ns.	1	0.9905	1
8	With unclassifiable category	Unclassifiable	Dinner	MBC:Cluster2	4012.27	0	0	11	1	ns.	1	0.9905	1
8	With unclassifiable category	Unclassifiable	Elf city	Intercept	-162.3016	4513.3762	NA	NA	NA	NA	1	0.9905	1
8	With unclassifiable category	Unclassifiable	Elf city	Valence	-1976.9952	24526.3461	0	11	1	ns.	1	0.9905	1
8	With unclassifiable category	Unclassifiable	Elf city	Arousal	-4789.9853	11917.372	0	11	1	ns.	1	0.9905	1
8	With unclassifiable category	Unclassifiable	Elf city	Dominance	-4097.4812	21529.8013	21.5599	11	0.028	*	1	0.9905	1
8	With unclassifiable category	Unclassifiable	Elf city	Uncertainty	168.2491	5437.9376	0	11	1	ns.	1	0.9905	1
8	With unclassifiable category	Unclassifiable	Elf city	MBC:Cluster2	2371.2013	4513.3762	0	11	1	ns.	1	0.9905	1
8	With unclassifiable category	Unclassifiable	Haunted House	Intercept	-15576.7717	0	NA	NA	NA	NA	1	0.9905	1
8	With unclassifiable category	Unclassifiable	Haunted House	Valence	4740.1153	0	0	11	1	ns.	1	0.9905	1
8	With unclassifiable category	Unclassifiable	Haunted House	Arousal	13746.4802	0	0	11	1	ns.	1	0.9905	1
8	With unclassifiable category	Unclassifiable	Haunted House	Dominance	-5620.7745	0	21.5599	11	0.028	*	1	0.9905	1
8	With unclassifiable category	Unclassifiable	Haunted House	Uncertainty	3363.7233	0	0	11	1	ns.	1	0.9905	1
8	With unclassifiable category	Unclassifiable	Haunted House	MBC:Cluster2	-450.062	0	0	11	1	ns.	1	0.9905	1
8	With unclassifiable category	Unclassifiable	sex	Intercept	-5449.2225	0	NA	NA	NA	NA	1	0.9905	1
8	With unclassifiable category	Unclassifiable	sex	Valence	-4815.4684	0	0	11	1	ns.	1	0.9905	1
8	With unclassifiable category	Unclassifiable	sex	Arousal	-3066.1975	0	0	11	1	ns.	1	0.9905	1
8	With unclassifiable category	Unclassifiable	sex	Dominance	-8214.2291	0	21.5599	11	0.028	*	1	0.9905	1
8	With unclassifiable category	Unclassifiable	sex	Uncertainty	-474.4479	0	0	11	1	ns.	1	0.9905	1
8	With unclassifiable category	Unclassifiable	sex	MBC:Cluster2	-5574.5795	0	0	11	1	ns.	1	0.9905	1
8	With unclassifiable category	Unclassifiable	Toy room	Intercept	-7303.1479	0	NA	NA	NA	NA	1	0.9905	1
8	With unclassifiable category	Unclassifiable	Toy room	Valence	2655.8426	0	0	11	1	ns.	1	0.9905	1
8	With unclassifiable category	Unclassifiable	Toy room	Arousal	9900.2167	0	0	11	1	ns.	1	0.9905	1
8	With unclassifiable category	Unclassifiable	Toy room	Dominance	6012.2062	0	21.5599	11	0.028	*	1	0.9905	1
8	With unclassifiable category	Unclassifiable	Toy room	Uncertainty	2042.8519	0	0	11	1	ns.	1	0.9905	1
8	With unclassifiable category	Unclassifiable	Toy room	MBC:Cluster2	1433.1895	0	0	11	1	ns.	1	0.9905	1
8	With unclassifiable category	Unclassifiable	Venice	Intercept	-4316.5398	0	NA	NA	NA	NA	1	0.9905	1
8	With unclassifiable category	Unclassifiable	Venice	Valence	12346.3083	0	0	11	1	ns.	1	0.9905	1
8	With unclassifiable category	Unclassifiable	Venice	Arousal	4430.712	0	0	11	1	ns.	1	0.9905	1
8	With unclassifiable category	Unclassifiable	Venice	Dominance	-3378.9132	0	21.5599	11	0.028	*	1	0.9905	1
8	With unclassifiable category	Unclassifiable	Venice	Uncertainty	1402.2895	0	0	11	1	ns.	1	0.9905	1
8	With unclassifiable category	Unclassifiable	Venice	MBC:Cluster2	-3996.9384	0	0	11	1	ns.	1	0.9905	1
8	With unclassifiable category	Unclassifiable	Village	Intercept	-2903.5727	0	NA	NA	NA	NA	1	0.9905	1
8	With unclassifiable category	Unclassifiable	Village	Valence	-4512.8896	NA	0	11	1	ns.	1	0.9905	1

ID	ModelType	Baseline	VECategory	TermInModel	Coefficient	StdErr	LRTest.ChiSq	DF	p	sig	McFadden	CoxSnell	Nagelkerke
8	With unclassifiable category	Unclassifiable	Village	Arousal	-7860.5184	NA	0	11	1	ns.	1	0.9905	1
8	With unclassifiable category	Unclassifiable	Village	Dominance	-10583.8622	0	21.5599	11	0.028	*	1	0.9905	1
8	With unclassifiable category	Unclassifiable	Village	Uncertainty	-395.2893	0	0	11	1	ns.	1	0.9905	1
8	With unclassifiable category	Unclassifiable	Village	MBC:Cluster2	4457.5857	NA	0	11	1	ns.	1	0.9905	1
9	Without unclassifiable category	relaxed	bored	Intercept	2649.7304	13535.0677	NA	NA	NA	NA	1	0.9674	1
9	Without unclassifiable category	relaxed	bored	Valence	-2956.2476	17567.751	30.1393	5	0	***	1	0.9674	1
9	Without unclassifiable category	relaxed	bored	Arousal	2890.6987	8772.4241	36.996	5	0	***	1	0.9674	1
9	Without unclassifiable category	relaxed	bored	Dominance	2622.5464	13899.4119	44.6981	5	0	***	1	0.9674	1
9	Without unclassifiable category	relaxed	bored	Uncertainty	-783.7738	3068.6236	24.5478	5	0.0002	***	1	0.9674	1
9	Without unclassifiable category	relaxed	bored	MBC:Cluster2	-672.0234	9290.0922	14.9752	5	0.0105	*	1	0.9674	1
9	Without unclassifiable category	relaxed	curious	Intercept	3010.7458	12137.321	NA	NA	NA	NA	1	0.9674	1
9	Without unclassifiable category	relaxed	curious	Valence	-1880.8125	26529.2996	30.1393	5	0	***	1	0.9674	1
9	Without unclassifiable category	relaxed	curious	Arousal	2455.8393	2682.1478	36.996	5	0	***	1	0.9674	1
9	Without unclassifiable category	relaxed	curious	Dominance	1617.9535	33084.3528	44.6981	5	0	***	1	0.9674	1
9	Without unclassifiable category	relaxed	curious	Uncertainty	-824.5651	2134.4563	24.5478	5	0.0002	***	1	0.9674	1
9	Without unclassifiable category	relaxed	curious	MBC:Cluster2	-1337.8242	11375.1324	14.9752	5	0.0105	*	1	0.9674	1
9	Without unclassifiable category	relaxed	sick	Intercept	-2284.7511	53498.2166	NA	NA	NA	NA	1	0.9674	1
9	Without unclassifiable category	relaxed	sick	Valence	414.7526	246703.7357	30.1393	5	0	***	1	0.9674	1
9	Without unclassifiable category	relaxed	sick	Arousal	5761.1375	3127.4415	36.996	5	0	***	1	0.9674	1
9	Without unclassifiable category	relaxed	sick	Dominance	264.7184	223956.9941	44.6981	5	0	***	1	0.9674	1
9	Without unclassifiable category	relaxed	sick	Uncertainty	-1088.5171	3946.2038	24.5478	5	0.0002	***	1	0.9674	1
9	Without unclassifiable category	relaxed	sick	MBC:Cluster2	5032.1376	53498.2166	14.9752	5	0.0105	*	1	0.9674	1
9	Without unclassifiable category	relaxed	spooky	Intercept	-3094.1141	101301.792	NA	NA	NA	NA	1	0.9674	1
9	Without unclassifiable category	relaxed	spooky	Valence	-3510.6845	216681.1433	30.1393	5	0	***	1	0.9674	1
9	Without unclassifiable category	relaxed	spooky	Arousal	-2212.2649	173088.1485	36.996	5	0	***	1	0.9674	1
9	Without unclassifiable category	relaxed	spooky	Dominance	-6197.0563	154407.4421	44.6981	5	0	***	1	0.9674	1
9	Without unclassifiable category	relaxed	spooky	Uncertainty	-7264.8534	15656.9703	24.5478	5	0.0002	***	1	0.9674	1
9	Without unclassifiable category	relaxed	spooky	MBC:Cluster2	443.3788	78682.2871	14.9752	5	0.0105	*	1	0.9674	1
9	Without unclassifiable category	relaxed	uncomfortable	Intercept	-1342.853	177242.0303	NA	NA	NA	NA	1	0.9674	1
9	Without unclassifiable category	relaxed	uncomfortable	Valence	-3089.2342	316415.3595	30.1393	5	0	***	1	0.9674	1
9	Without unclassifiable category	relaxed	uncomfortable	Arousal	308.778	35473.6699	36.996	5	0	***	1	0.9674	1
9	Without unclassifiable category	relaxed	uncomfortable	Dominance	-2814.6012	163762.7074	44.6981	5	0	***	1	0.9674	1
9	Without unclassifiable category	relaxed	uncomfortable	Uncertainty	-6037.9996	40497.6051	24.5478	5	0.0002	***	1	0.9674	1
9	Without unclassifiable category	relaxed	uncomfortable	MBC:Cluster2	-2198.1064	0	14.9752	5	0.0105	*	1	0.9674	1
10	Without unclassifiable category	fantasylike	sex	Intercept	-8.3178	49.8333	NA	NA	NA	NA	0.5508	0.8192	0.8576
10	Without unclassifiable category	fantasylike	sex	Valence	1.3771	3.3735	4.2222	4	0.3768	ns.	0.5508	0.8192	0.8576
10	Without unclassifiable category	fantasylike	sex	Arousal	-11.7661	9.7081	4.5698	4	0.3344	ns.	0.5508	0.8192	0.8576
10	Without unclassifiable category	fantasylike	sex	Dominance	-12.8836	12.4787	3.9818	4	0.4085	ns.	0.5508	0.8192	0.8576
10	Without unclassifiable category	fantasylike	sex	Uncertainty	-43.1311	215.8765	5.2147	4	0.266	ns.	0.5508	0.8192	0.8576
10	Without unclassifiable category	fantasylike	sex	MBC:Cluster2	-40.611	0	6.3295	4	0.1759	ns.	0.5508	0.8192	0.8576
10	Without unclassifiable category	fantasylike	fpslike	Intercept	0.4807	1.9018	NA	NA	NA	NA	0.5508	0.8192	0.8576
10	Without unclassifiable category	fantasylike	fpslike	Valence	-1.4569	2.2671	4.2222	4	0.3768	ns.	0.5508	0.8192	0.8576
10	Without unclassifiable category	fantasylike	fpslike	Arousal	-0.5968	1.6527	4.5698	4	0.3344	ns.	0.5508	0.8192	0.8576
10	Without unclassifiable category	fantasylike	fpslike	Dominance	1.2676	2.2808	3.9818	4	0.4085	ns.	0.5508	0.8192	0.8576
10	Without unclassifiable category	fantasylike	fpslike	Uncertainty	0.4657	1.0437	5.2147	4	0.266	ns.	0.5508	0.8192	0.8576
10	Without unclassifiable category	fantasylike	fpslike	MBC:Cluster2	-1.9589	3.124	6.3295	4	0.1759	ns.	0.5508	0.8192	0.8576
10	Without unclassifiable category	fantasylike	outdoor	Intercept	-32.8469	1033.3968	NA	NA	NA	NA	0.5508	0.8192	0.8576
10	Without unclassifiable category	fantasylike	outdoor	Valence	39.3104	441.0188	4.2222	4	0.3768	ns.	0.5508	0.8192	0.8576
10	Without unclassifiable category	fantasylike	outdoor	Arousal	-2.4994	1418.4429	4.5698	4	0.3344	ns.	0.5508	0.8192	0.8576
10	Without unclassifiable category	fantasylike	outdoor	Dominance	22.4614	550.7727	3.9818	4	0.4085	ns.	0.5508	0.8192	0.8576
10	Without unclassifiable category	fantasylike	outdoor	Uncertainty	6.686	370.4867	5.2147	4	0.266	ns.	0.5508	0.8192	0.8576
10	Without unclassifiable category	fantasylike	outdoor	MBC:Cluster2	-38.5871	1035.1102	6.3295	4	0.1759	ns.	0.5508	0.8192	0.8576

ID	ModelType	Baseline	VECategory	TermInModel	Coefficient	StdErr	LRTest.ChiSq	DF	p	sig	McFadden	CoxSnell	Nagelkerke
10	Without unclassifiable category	fantasylike	realworld	Intercept	-30.7241	128.1628	NA	NA	NA	NA	0.5508	0.8192	0.8576
10	Without unclassifiable category	fantasylike	realworld	Valence	-1.0167	4.3735	4.2222	4	0.3768	ns.	0.5508	0.8192	0.8576
10	Without unclassifiable category	fantasylike	realworld	Arousal	0.7237	3.0806	4.5698	4	0.3344	ns.	0.5508	0.8192	0.8576
10	Without unclassifiable category	fantasylike	realworld	Dominance	0.7292	2.0492	3.9818	4	0.4085	ns.	0.5508	0.8192	0.8576
10	Without unclassifiable category	fantasylike	realworld	Uncertainty	5.4915	25.4347	5.2147	4	0.266	ns.	0.5508	0.8192	0.8576
10	Without unclassifiable category	fantasylike	realworld	MBC:Cluster2	33.0081	131.82	6.3295	4	0.1759	ns.	0.5508	0.8192	0.8576
11	With unclassifiable category	Unclassifiable	scary	Intercept	-33.9377	1105.4588	NA	NA	NA	NA	0.9044	0.956	0.9872
11	With unclassifiable category	Unclassifiable	scary	Valence	18.3192	722.7758	2.7312	6	0.8417	ns.	0.9044	0.956	0.9872
11	With unclassifiable category	Unclassifiable	scary	Arousal	193.9845	519.2708	18.1144	6	0.006	**	0.9044	0.956	0.9872
11	With unclassifiable category	Unclassifiable	scary	Dominance	152.1703	593.3803	11.3887	6	0.0771	ns.	0.9044	0.956	0.9872
11	With unclassifiable category	Unclassifiable	scary	Uncertainty	-36.998	253.2781	9.6336	6	0.141	ns.	0.9044	0.956	0.9872
11	With unclassifiable category	Unclassifiable	scary	MBC:Cluster2	-54.4462	0.0053	6.3042	6	0.39	ns.	0.9044	0.956	0.9872
11	With unclassifiable category	Unclassifiable	relaxed	Intercept	-21.1547	64.9195	NA	NA	NA	NA	0.9044	0.956	0.9872
11	With unclassifiable category	Unclassifiable	relaxed	Valence	62.1962	297.1547	2.7312	6	0.8417	ns.	0.9044	0.956	0.9872
11	With unclassifiable category	Unclassifiable	relaxed	Arousal	-76.1206	162.8707	18.1144	6	0.006	**	0.9044	0.956	0.9872
11	With unclassifiable category	Unclassifiable	relaxed	Dominance	36.6436	326.3695	11.3887	6	0.0771	ns.	0.9044	0.956	0.9872
11	With unclassifiable category	Unclassifiable	relaxed	Uncertainty	-15.7561	292.6502	9.6336	6	0.141	ns.	0.9044	0.956	0.9872
11	With unclassifiable category	Unclassifiable	relaxed	MBC:Cluster2	-37.2586	64.9195	6.3042	6	0.39	ns.	0.9044	0.956	0.9872
11	With unclassifiable category	Unclassifiable	Childlike	Intercept	-87.162	0.2029	NA	NA	NA	NA	0.9044	0.956	0.9872
11	With unclassifiable category	Unclassifiable	Childlike	Valence	-134.6735	0.2206	2.7312	6	0.8417	ns.	0.9044	0.956	0.9872
11	With unclassifiable category	Unclassifiable	Childlike	Arousal	55.281	0.0687	18.1144	6	0.006	**	0.9044	0.956	0.9872
11	With unclassifiable category	Unclassifiable	Childlike	Dominance	210.4378	0.2552	11.3887	6	0.0771	ns.	0.9044	0.956	0.9872
11	With unclassifiable category	Unclassifiable	Childlike	Uncertainty	-15.8743	0.0465	9.6336	6	0.141	ns.	0.9044	0.956	0.9872
11	With unclassifiable category	Unclassifiable	Childlike	MBC:Cluster2	71.6039	0.2184	6.3042	6	0.39	ns.	0.9044	0.956	0.9872
11	With unclassifiable category	Unclassifiable	intrigued	Intercept	-56.2129	124.9314	NA	NA	NA	NA	0.9044	0.956	0.9872
11	With unclassifiable category	Unclassifiable	intrigued	Valence	176.8229	142.1794	2.7312	6	0.8417	ns.	0.9044	0.956	0.9872
11	With unclassifiable category	Unclassifiable	intrigued	Arousal	168.927	425.2917	18.1144	6	0.006	**	0.9044	0.956	0.9872
11	With unclassifiable category	Unclassifiable	intrigued	Dominance	-19.1713	147.0888	11.3887	6	0.0771	ns.	0.9044	0.956	0.9872
11	With unclassifiable category	Unclassifiable	intrigued	Uncertainty	-29.2828	28.6433	9.6336	6	0.141	ns.	0.9044	0.956	0.9872
11	With unclassifiable category	Unclassifiable	intrigued	MBC:Cluster2	7.9652	0	6.3042	6	0.39	ns.	0.9044	0.956	0.9872
11	With unclassifiable category	Unclassifiable	macho	Intercept	-62.6802	862.5889	NA	NA	NA	NA	0.9044	0.956	0.9872
11	With unclassifiable category	Unclassifiable	macho	Valence	-1.66	137.1624	2.7312	6	0.8417	ns.	0.9044	0.956	0.9872
11	With unclassifiable category	Unclassifiable	macho	Arousal	-29.0659	734.067	18.1144	6	0.006	**	0.9044	0.956	0.9872
11	With unclassifiable category	Unclassifiable	macho	Dominance	-89.313	176.0664	11.3887	6	0.0771	ns.	0.9044	0.956	0.9872
11	With unclassifiable category	Unclassifiable	macho	Uncertainty	-78.5013	195.9692	9.6336	6	0.141	ns.	0.9044	0.956	0.9872
11	With unclassifiable category	Unclassifiable	macho	MBC:Cluster2	-72.2843	NA	6.3042	6	0.39	ns.	0.9044	0.956	0.9872
11	With unclassifiable category	Unclassifiable	uncanny and empty	Intercept	157.9889	793.8147	NA	NA	NA	NA	0.9044	0.956	0.9872
11	With unclassifiable category	Unclassifiable	uncanny and empty	Valence	35.8517	297.1722	2.7312	6	0.8417	ns.	0.9044	0.956	0.9872
11	With unclassifiable category	Unclassifiable	uncanny and empty	Arousal	-73.2477	162.4994	18.1144	6	0.006	**	0.9044	0.956	0.9872
11	With unclassifiable category	Unclassifiable	uncanny and empty	Dominance	47.961	326.484	11.3887	6	0.0771	ns.	0.9044	0.956	0.9872
11	With unclassifiable category	Unclassifiable	uncanny and empty	Uncertainty	-60.9527	262.3166	9.6336	6	0.141	ns.	0.9044	0.956	0.9872
11	With unclassifiable category	Unclassifiable	uncanny and empty	MBC:Cluster2	-210.8046	916.3046	6.3042	6	0.39	ns.	0.9044	0.956	0.9872
12	Without unclassifiable category	spooky	Childlike	Intercept	-256.4419	116.8997	NA	NA	NA	NA	0.9258	0.9591	0.9904
12	Without unclassifiable category	spooky	Childlike	Valence	318.4098	142.3944	1.4965	5	0.9135	ns.	0.9258	0.9591	0.9904
12	Without unclassifiable category	spooky	Childlike	Arousal	438.9394	444.8655	20.7943	5	0.0009	***	0.9258	0.9591	0.9904
12	Without unclassifiable category	spooky	Childlike	Dominance	468.6277	192.7574	18.3366	5	0.0026	**	0.9258	0.9591	0.9904

ID	ModelType	Baseline	VECategory	TermInModel	Coefficient	StdErr	LRTest.ChiSq	DF	p	sig	McFadden	CoxSnell	Nagelkerke
12	Without unclassifiable category	spooky	Childlike	Uncertainty	-100.3102	177.0822	12.3831	5	0.0299	*	0.9258	0.9591	0.9904
12	Without unclassifiable category	spooky	Childlike	MBC:Cluster2	380.4317	116.8997	6.3325	5	0.2752	ns.	0.9258	0.9591	0.9904
12	Without unclassifiable category	spooky	enchanted	Intercept	-63.4071	111.4128	NA	NA	NA	NA	0.9258	0.9591	0.9904
12	Without unclassifiable category	spooky	enchanted	Valence	298.7578	152.3452	1.4965	5	0.9135	ns.	0.9258	0.9591	0.9904
12	Without unclassifiable category	spooky	enchanted	Arousal	-160.999	352.6102	20.7943	5	0.0009	***	0.9258	0.9591	0.9904
12	Without unclassifiable category	spooky	enchanted	Dominance	374.2848	190.6207	18.3366	5	0.0026	**	0.9258	0.9591	0.9904
12	Without unclassifiable category	spooky	enchanted	Uncertainty	-150.2728	184.0345	12.3831	5	0.0299	*	0.9258	0.9591	0.9904
12	Without unclassifiable category	spooky	enchanted	MBC:Cluster2	-130.9082	111.4128	6.3325	5	0.2752	ns.	0.9258	0.9591	0.9904
12	Without unclassifiable category	spooky	escorts	Intercept	-106.4425	5.0888	NA	NA	NA	NA	0.9258	0.9591	0.9904
12	Without unclassifiable category	spooky	escorts	Valence	2.172	3.1971	1.4965	5	0.9135	ns.	0.9258	0.9591	0.9904
12	Without unclassifiable category	spooky	escorts	Arousal	-11.969	10.6507	20.7943	5	0.0009	***	0.9258	0.9591	0.9904
12	Without unclassifiable category	spooky	escorts	Dominance	-14.133	13.6504	18.3366	5	0.0026	**	0.9258	0.9591	0.9904
12	Without unclassifiable category	spooky	escorts	Uncertainty	-466.963	1.1686	12.3831	5	0.0299	*	0.9258	0.9591	0.9904
12	Without unclassifiable category	spooky	escorts	MBC:Cluster2	-337.155	0	6.3325	5	0.2752	ns.	0.9258	0.9591	0.9904
12	Without unclassifiable category	spooky	interior design	Intercept	313.738	309.9564	NA	NA	NA	NA	0.9258	0.9591	0.9904
12	Without unclassifiable category	spooky	interior design	Valence	108.1143	121.0889	1.4965	5	0.9135	ns.	0.9258	0.9591	0.9904
12	Without unclassifiable category	spooky	interior design	Arousal	-224.2867	316.4009	20.7943	5	0.0009	***	0.9258	0.9591	0.9904
12	Without unclassifiable category	spooky	interior design	Dominance	263.0787	486.8779	18.3366	5	0.0026	**	0.9258	0.9591	0.9904
12	Without unclassifiable category	spooky	interior design	Uncertainty	-114.0913	99.4637	12.3831	5	0.0299	*	0.9258	0.9591	0.9904
12	Without unclassifiable category	spooky	interior design	MBC:Cluster2	-464.3202	218.474	6.3325	5	0.2752	ns.	0.9258	0.9591	0.9904
12	Without unclassifiable category	spooky	water	Intercept	-149.5258	192.3742	NA	NA	NA	NA	0.9258	0.9591	0.9904
12	Without unclassifiable category	spooky	water	Valence	374.5422	84.1123	1.4965	5	0.9135	ns.	0.9258	0.9591	0.9904
12	Without unclassifiable category	spooky	water	Arousal	262.4761	112.1353	20.7943	5	0.0009	***	0.9258	0.9591	0.9904
12	Without unclassifiable category	spooky	water	Dominance	759.5598	236.6499	18.3366	5	0.0026	**	0.9258	0.9591	0.9904
12	Without unclassifiable category	spooky	water	Uncertainty	-63.4264	43.3504	12.3831	5	0.0299	*	0.9258	0.9591	0.9904
12	Without unclassifiable category	spooky	water	MBC:Cluster2	-189.5645	192.3742	6.3325	5	0.2752	ns.	0.9258	0.9591	0.9904
13	With unclassifiable category	Unclassifiable	happy	Intercept	-5.1018	208.836	NA	NA	NA	NA	0.7672	0.8571	0.9308
13	With unclassifiable category	Unclassifiable	happy	Valence	-8.8516	179.0353	4.2677	3	0.234	ns.	0.7672	0.8571	0.9308
13	With unclassifiable category	Unclassifiable	happy	Arousal	42.3976	382.5049	0.2929	3	0.9614	ns.	0.7672	0.8571	0.9308
13	With unclassifiable category	Unclassifiable	happy	Dominance	42.9458	401.5702	9.7581	3	0.0207	*	0.7672	0.8571	0.9308
13	With unclassifiable category	Unclassifiable	happy	Uncertainty	-32.5759	148.8852	4.0739	3	0.2536	ns.	0.7672	0.8571	0.9308
13	With unclassifiable category	Unclassifiable	happy	MBC:Cluster2	29.04	304.2734	3.6594	3	0.3007	ns.	0.7672	0.8571	0.9308
13	With unclassifiable category	Unclassifiable	distressing	Intercept	38.6402	155.3063	NA	NA	NA	NA	0.7672	0.8571	0.9308
13	With unclassifiable category	Unclassifiable	distressing	Valence	65.0289	173.6331	4.2677	3	0.234	ns.	0.7672	0.8571	0.9308
13	With unclassifiable category	Unclassifiable	distressing	Arousal	43.7145	353.6326	0.2929	3	0.9614	ns.	0.7672	0.8571	0.9308
13	With unclassifiable category	Unclassifiable	distressing	Dominance	-39.6503	387.7641	9.7581	3	0.0207	*	0.7672	0.8571	0.9308
13	With unclassifiable category	Unclassifiable	distressing	Uncertainty	-38.3015	83.531	4.0739	3	0.2536	ns.	0.7672	0.8571	0.9308
13	With unclassifiable category	Unclassifiable	distressing	MBC:Cluster2	-35.4194	205.515	3.6594	3	0.3007	ns.	0.7672	0.8571	0.9308
13	With unclassifiable category	Unclassifiable	dull	Intercept	39.029	155.3422	NA	NA	NA	NA	0.7672	0.8571	0.9308
13	With unclassifiable category	Unclassifiable	dull	Valence	65.2053	173.6305	4.2677	3	0.234	ns.	0.7672	0.8571	0.9308
13	With unclassifiable category	Unclassifiable	dull	Arousal	43.458	353.6221	0.2929	3	0.9614	ns.	0.7672	0.8571	0.9308
13	With unclassifiable category	Unclassifiable	dull	Dominance	-37.2751	387.7553	9.7581	3	0.0207	*	0.7672	0.8571	0.9308
13	With unclassifiable category	Unclassifiable	dull	Uncertainty	-43.357	83.579	4.0739	3	0.2536	ns.	0.7672	0.8571	0.9308
13	With unclassifiable category	Unclassifiable	dull	MBC:Cluster2	-36.2979	205.535	3.6594	3	0.3007	ns.	0.7672	0.8571	0.9308
14	Without unclassifiable category	Nice garden	scary	Intercept	-3343.5602	0	NA	NA	NA	NA	1	0.9914	1
14	Without unclassifiable category	Nice garden	scary	Valence	559.4474	0	0	12	1	ns.	1	0.9914	1

ID	ModelType	Baseline	VECategory	TermInModel	Coefficient	StdErr	LRTest.ChiSq	DF	p	sig	McFadden	CoxSnell	Nagelkerke
14	Without unclassifiable category	Nice garden	scary	Arousal	4779.6909	0	4.1391	12	0.9808	ns.	1	0.9914	1
14	Without unclassifiable category	Nice garden	scary	Dominance	-2087.9966	0	10.0703	12	0.6098	ns.	1	0.9914	1
14	Without unclassifiable category	Nice garden	scary	Uncertainty	632.8271	0	-0.0001	12	1	ns.	1	0.9914	1
14	Without unclassifiable category	Nice garden	scary	MBC:Cluster2	-147.4785	0	-0.0001	12	1	ns.	1	0.9914	1
14	Without unclassifiable category	Nice garden	Childish buildings	Intercept	-357.0104	0.0001	NA	NA	NA	NA	1	0.9914	1
14	Without unclassifiable category	Nice garden	Childish buildings	Valence	-2900.1116	0.0001	0	12	1	ns.	1	0.9914	1
14	Without unclassifiable category	Nice garden	Childish buildings	Arousal	4273.8718	0	4.1391	12	0.9808	ns.	1	0.9914	1
14	Without unclassifiable category	Nice garden	Childish buildings	Dominance	4206.7089	0	10.0703	12	0.6098	ns.	1	0.9914	1
14	Without unclassifiable category	Nice garden	Childish buildings	Uncertainty	61.705	NA	-0.0001	12	1	ns.	1	0.9914	1
14	Without unclassifiable category	Nice garden	Childish buildings	MBC:Cluster2	680.7084	NA	-0.0001	12	1	ns.	1	0.9914	1
14	Without unclassifiable category	Nice garden	Church	Intercept	-1137.5858	0	NA	NA	NA	NA	1	0.9914	1
14	Without unclassifiable category	Nice garden	Church	Valence	220.5806	NA	0	12	1	ns.	1	0.9914	1
14	Without unclassifiable category	Nice garden	Church	Arousal	-2189.1471	0	4.1391	12	0.9808	ns.	1	0.9914	1
14	Without unclassifiable category	Nice garden	Church	Dominance	-4575.3042	0	10.0703	12	0.6098	ns.	1	0.9914	1
14	Without unclassifiable category	Nice garden	Church	Uncertainty	-79.1817	0	-0.0001	12	1	ns.	1	0.9914	1
14	Without unclassifiable category	Nice garden	Church	MBC:Cluster2	-660.3777	0	-0.0001	12	1	ns.	1	0.9914	1
14	Without unclassifiable category	Nice garden	Clean buildings	Intercept	-1547.3484	NA	NA	NA	NA	NA	1	0.9914	1
14	Without unclassifiable category	Nice garden	Clean buildings	Valence	-3849.566	0	0	12	1	ns.	1	0.9914	1
14	Without unclassifiable category	Nice garden	Clean buildings	Arousal	-1521.7185	0	4.1391	12	0.9808	ns.	1	0.9914	1
14	Without unclassifiable category	Nice garden	Clean buildings	Dominance	-2844.4689	0	10.0703	12	0.6098	ns.	1	0.9914	1
14	Without unclassifiable category	Nice garden	Clean buildings	Uncertainty	351.0456	0	-0.0001	12	1	ns.	1	0.9914	1
14	Without unclassifiable category	Nice garden	Clean buildings	MBC:Cluster2	2007.6061	0	-0.0001	12	1	ns.	1	0.9914	1
14	Without unclassifiable category	Nice garden	Crude buildings	Intercept	-330.9937	19579.5693	NA	NA	NA	NA	1	0.9914	1
14	Without unclassifiable category	Nice garden	Crude buildings	Valence	-2189.8322	180172.1007	0	12	1	ns.	1	0.9914	1
14	Without unclassifiable category	Nice garden	Crude buildings	Arousal	-422.7627	111134.4118	4.1391	12	0.9808	ns.	1	0.9914	1
14	Without unclassifiable category	Nice garden	Crude buildings	Dominance	-3481.5017	59680.3667	10.0703	12	0.6098	ns.	1	0.9914	1
14	Without unclassifiable category	Nice garden	Crude buildings	Uncertainty	331.8087	4430.127	-0.0001	12	1	ns.	1	0.9914	1
14	Without unclassifiable category	Nice garden	Crude buildings	MBC:Cluster2	-2960.4035	0	-0.0001	12	1	ns.	1	0.9914	1
14	Without unclassifiable category	Nice garden	Dining	Intercept	-57.7858	0	NA	NA	NA	NA	1	0.9914	1
14	Without unclassifiable category	Nice garden	Dining	Valence	1684.3661	0	0	12	1	ns.	1	0.9914	1
14	Without unclassifiable category	Nice garden	Dining	Arousal	1815.9815	NA	4.1391	12	0.9808	ns.	1	0.9914	1
14	Without unclassifiable category	Nice garden	Dining	Dominance	-697.2347	0	10.0703	12	0.6098	ns.	1	0.9914	1
14	Without unclassifiable category	Nice garden	Dining	Uncertainty	-97.2848	NA	-0.0001	12	1	ns.	1	0.9914	1
14	Without unclassifiable category	Nice garden	Dining	MBC:Cluster2	225.5497	0	-0.0001	12	1	ns.	1	0.9914	1

ID	ModelType	Baseline	VECategory	TermInModel	Coefficient	StdErr	LRTest.ChiSq	DF	p	sig	McFadden	CoxSnell	Nagelkerke
14	Without unclassifiable category	Nice garden	Dirty beach	Intercept	-706.3427	NA	NA	NA	NA	NA	1	0.9914	1
14	Without unclassifiable category	Nice garden	Dirty beach	Valence	-4413.4451	NA	0	12	1	ns.	1	0.9914	1
14	Without unclassifiable category	Nice garden	Dirty beach	Arousal	799.5061	NA	4.1391	12	0.9808	ns.	1	0.9914	1
14	Without unclassifiable category	Nice garden	Dirty beach	Dominance	983.6153	NA	10.0703	12	0.6098	ns.	1	0.9914	1
14	Without unclassifiable category	Nice garden	Dirty beach	Uncertainty	-193.3066	NA	-0.0001	12	1	ns.	1	0.9914	1
14	Without unclassifiable category	Nice garden	Dirty beach	MBC:Cluster2	-1137.8159	0	-0.0001	12	1	ns.	1	0.9914	1
14	Without unclassifiable category	Nice garden	Dirty building	Intercept	947.1913	19579.5693	NA	NA	NA	NA	1	0.9914	1
14	Without unclassifiable category	Nice garden	Dirty building	Valence	-1871.143	180172.1004	0	12	1	ns.	1	0.9914	1
14	Without unclassifiable category	Nice garden	Dirty building	Arousal	-74.3854	111134.4117	4.1391	12	0.9808	ns.	1	0.9914	1
14	Without unclassifiable category	Nice garden	Dirty building	Dominance	-1970.0455	59680.3667	10.0703	12	0.6098	ns.	1	0.9914	1
14	Without unclassifiable category	Nice garden	Dirty building	Uncertainty	-595.3586	4430.127	-0.0001	12	1	ns.	1	0.9914	1
14	Without unclassifiable category	Nice garden	Dirty building	MBC:Cluster2	-1311.6112	0	-0.0001	12	1	ns.	1	0.9914	1
14	Without unclassifiable category	Nice garden	Modern Venice	Intercept	-1231.1505	0	NA	NA	NA	NA	1	0.9914	1
14	Without unclassifiable category	Nice garden	Modern Venice	Valence	3789.5505	NA	0	12	1	ns.	1	0.9914	1
14	Without unclassifiable category	Nice garden	Modern Venice	Arousal	2136.6148	NA	4.1391	12	0.9808	ns.	1	0.9914	1
14	Without unclassifiable category	Nice garden	Modern Venice	Dominance	-367.7817	NA	10.0703	12	0.6098	ns.	1	0.9914	1
14	Without unclassifiable category	Nice garden	Modern Venice	Uncertainty	338.2719	NA	-0.0001	12	1	ns.	1	0.9914	1
14	Without unclassifiable category	Nice garden	Modern Venice	MBC:Cluster2	-1312.8575	0	-0.0001	12	1	ns.	1	0.9914	1
14	Without unclassifiable category	Nice garden	Nice beach	Intercept	-1573.1538	0	NA	NA	NA	NA	1	0.9914	1
14	Without unclassifiable category	Nice garden	Nice beach	Valence	1068.6583	0	0	12	1	ns.	1	0.9914	1
14	Without unclassifiable category	Nice garden	Nice beach	Arousal	3104.4176	0	4.1391	12	0.9808	ns.	1	0.9914	1
14	Without unclassifiable category	Nice garden	Nice beach	Dominance	3368.1402	0	10.0703	12	0.6098	ns.	1	0.9914	1
14	Without unclassifiable category	Nice garden	Nice beach	Uncertainty	403.8636	0	-0.0001	12	1	ns.	1	0.9914	1
14	Without unclassifiable category	Nice garden	Nice beach	MBC:Cluster2	-1659.5102	0	-0.0001	12	1	ns.	1	0.9914	1
14	Without unclassifiable category	Nice garden	Themepark	Intercept	366.5216	50499.5279	NA	NA	NA	NA	1	0.9914	1
14	Without unclassifiable category	Nice garden	Themepark	Valence	-498.9922	208443.1208	0	12	1	ns.	1	0.9914	1
14	Without unclassifiable category	Nice garden	Themepark	Arousal	1344.7422	91491.6322	4.1391	12	0.9808	ns.	1	0.9914	1
14	Without unclassifiable category	Nice garden	Themepark	Dominance	555.5515	145584.6079	10.0703	12	0.6098	ns.	1	0.9914	1
14	Without unclassifiable category	Nice garden	Themepark	Uncertainty	-116.378	45540.2511	-0.0001	12	1	ns.	1	0.9914	1
14	Without unclassifiable category	Nice garden	Themepark	MBC:Cluster2	463.1002	50499.5279	-0.0001	12	1	ns.	1	0.9914	1
14	Without unclassifiable category	Nice garden	Weird buildings	Intercept	1724.1616	0	NA	NA	NA	NA	1	0.9914	1
14	Without unclassifiable category	Nice garden	Weird buildings	Valence	555.1306	0	0	12	1	ns.	1	0.9914	1
14	Without unclassifiable category	Nice garden	Weird buildings	Arousal	3386.7473	0	4.1391	12	0.9808	ns.	1	0.9914	1
14	Without unclassifiable category	Nice garden	Weird buildings	Dominance	1321.731	0	10.0703	12	0.6098	ns.	1	0.9914	1
14	Without unclassifiable category	Nice garden	Weird buildings	Uncertainty	757.3309	0	-0.0001	12	1	ns.	1	0.9914	1

ID	ModelType	Baseline	VECategory	TermInModel	Coefficient	StdErr	LRTest.ChiSq	DF	p	sig	McFadden	CoxSnell	Nagelkerke
14	Without unclassifiable category	Nice garden	Weird buildings	MBC:Cluster2	-2096.7294	0	-0.0001	12	1	ns.	1	0.9914	1
15	With unclassifiable category	Unclassifiable	scary	Intercept	2.4633	67.517	NA	NA	NA	NA	0.8377	0.9644	0.9827
15	With unclassifiable category	Unclassifiable	scary	Valence	-15.6237	87.2227	9.6104	7	0.2117	ns.	0.8377	0.9644	0.9827
15	With unclassifiable category	Unclassifiable	scary	Arousal	-37.1066	69.1012	5.608	7	0.5862	ns.	0.8377	0.9644	0.9827
15	With unclassifiable category	Unclassifiable	scary	Dominance	-151.4421	42.3238	11.47	7	0.1194	ns.	0.8377	0.9644	0.9827
15	With unclassifiable category	Unclassifiable	scary	Uncertainty	119.1891	11.5246	5.2027	7	0.6352	ns.	0.8377	0.9644	0.9827
15	With unclassifiable category	Unclassifiable	scary	MBC:Cluster2	-25.3501	101.0964	6.1566	7	0.5216	ns.	0.8377	0.9644	0.9827
15	With unclassifiable category	Unclassifiable	relaxed	Intercept	-11.532	123.6922	NA	NA	NA	NA	0.8377	0.9644	0.9827
15	With unclassifiable category	Unclassifiable	relaxed	Valence	-86.7393	97.9575	9.6104	7	0.2117	ns.	0.8377	0.9644	0.9827
15	With unclassifiable category	Unclassifiable	relaxed	Arousal	-122.6148	170.6768	5.608	7	0.5862	ns.	0.8377	0.9644	0.9827
15	With unclassifiable category	Unclassifiable	relaxed	Dominance	80.7071	118.2985	11.47	7	0.1194	ns.	0.8377	0.9644	0.9827
15	With unclassifiable category	Unclassifiable	relaxed	Uncertainty	-11.67	709.3958	5.2027	7	0.6352	ns.	0.8377	0.9644	0.9827
15	With unclassifiable category	Unclassifiable	relaxed	MBC:Cluster2	-42.0346	123.6922	6.1566	7	0.5216	ns.	0.8377	0.9644	0.9827
15	With unclassifiable category	Unclassifiable	empty	Intercept	-50.9502	0.0732	NA	NA	NA	NA	0.8377	0.9644	0.9827
15	With unclassifiable category	Unclassifiable	empty	Valence	-162.1242	0.0008	9.6104	7	0.2117	ns.	0.8377	0.9644	0.9827
15	With unclassifiable category	Unclassifiable	empty	Arousal	-128.5908	0.052	5.608	7	0.5862	ns.	0.8377	0.9644	0.9827
15	With unclassifiable category	Unclassifiable	empty	Dominance	88.6092	0.0001	11.47	7	0.1194	ns.	0.8377	0.9644	0.9827
15	With unclassifiable category	Unclassifiable	empty	Uncertainty	138.8916	0.0063	5.2027	7	0.6352	ns.	0.8377	0.9644	0.9827
15	With unclassifiable category	Unclassifiable	empty	MBC:Cluster2	27.8493	0.0732	6.1566	7	0.5216	ns.	0.8377	0.9644	0.9827
15	With unclassifiable category	Unclassifiable	Childhood	Intercept	-86.3398	62.4357	NA	NA	NA	NA	0.8377	0.9644	0.9827
15	With unclassifiable category	Unclassifiable	Childhood	Valence	-92.6149	97.1086	9.6104	7	0.2117	ns.	0.8377	0.9644	0.9827
15	With unclassifiable category	Unclassifiable	Childhood	Arousal	-82.3006	175.2442	5.608	7	0.5862	ns.	0.8377	0.9644	0.9827
15	With unclassifiable category	Unclassifiable	Childhood	Dominance	81.0142	117.3291	11.47	7	0.1194	ns.	0.8377	0.9644	0.9827
15	With unclassifiable category	Unclassifiable	Childhood	Uncertainty	-57.0807	913.1745	5.2027	7	0.6352	ns.	0.8377	0.9644	0.9827
15	With unclassifiable category	Unclassifiable	Childhood	MBC:Cluster2	53.7579	62.4357	6.1566	7	0.5216	ns.	0.8377	0.9644	0.9827
15	With unclassifiable category	Unclassifiable	Dirty	Intercept	-45.3236	119.0678	NA	NA	NA	NA	0.8377	0.9644	0.9827
15	With unclassifiable category	Unclassifiable	Dirty	Valence	-147.6606	97.3725	9.6104	7	0.2117	ns.	0.8377	0.9644	0.9827
15	With unclassifiable category	Unclassifiable	Dirty	Arousal	72.1933	103.0553	5.608	7	0.5862	ns.	0.8377	0.9644	0.9827
15	With unclassifiable category	Unclassifiable	Dirty	Dominance	168.4735	41.3768	11.47	7	0.1194	ns.	0.8377	0.9644	0.9827
15	With unclassifiable category	Unclassifiable	Dirty	Uncertainty	-13.5381	27.2521	5.2027	7	0.6352	ns.	0.8377	0.9644	0.9827
15	With unclassifiable category	Unclassifiable	Dirty	MBC:Cluster2	-6.3433	2.6262	6.1566	7	0.5216	ns.	0.8377	0.9644	0.9827
15	With unclassifiable category	Unclassifiable	Luxury	Intercept	108.6747	139.1439	NA	NA	NA	NA	0.8377	0.9644	0.9827
15	With unclassifiable category	Unclassifiable	Luxury	Valence	-66.0995	93.7649	9.6104	7	0.2117	ns.	0.8377	0.9644	0.9827
15	With unclassifiable category	Unclassifiable	Luxury	Arousal	-86.275	171.3622	5.608	7	0.5862	ns.	0.8377	0.9644	0.9827
15	With unclassifiable category	Unclassifiable	Luxury	Dominance	63.2678	116.938	11.47	7	0.1194	ns.	0.8377	0.9644	0.9827
15	With unclassifiable category	Unclassifiable	Luxury	Uncertainty	44.3037	833.6032	5.2027	7	0.6352	ns.	0.8377	0.9644	0.9827
15	With unclassifiable category	Unclassifiable	Luxury	MBC:Cluster2	-128.8419	73.6708	6.1566	7	0.5216	ns.	0.8377	0.9644	0.9827
15	With unclassifiable category	Unclassifiable	x-rated	Intercept	-57.8229	66.5102	NA	NA	NA	NA	0.8377	0.9644	0.9827
15	With unclassifiable category	Unclassifiable	x-rated	Valence	-14.2962	87.2305	9.6104	7	0.2117	ns.	0.8377	0.9644	0.9827
15	With unclassifiable category	Unclassifiable	x-rated	Arousal	-46.1197	69.1633	5.608	7	0.5862	ns.	0.8377	0.9644	0.9827
15	With unclassifiable category	Unclassifiable	x-rated	Dominance	-161.2031	42.323	11.47	7	0.1194	ns.	0.8377	0.9644	0.9827
15	With unclassifiable category	Unclassifiable	x-rated	Uncertainty	-149.2049	15.2497	5.2027	7	0.6352	ns.	0.8377	0.9644	0.9827
15	With unclassifiable category	Unclassifiable	x-rated	MBC:Cluster2	-58.1156	0	6.1566	7	0.5216	ns.	0.8377	0.9644	0.9827
16	Without unclassifiable category	unhappy	scary	Intercept	-44.0026	75.406	NA	NA	NA	NA	0.808	0.9397	0.9697
16	Without unclassifiable category	unhappy	scary	Valence	29.7813	178.6116	3.4963	5	0.6239	ns.	0.808	0.9397	0.9697
16	Without unclassifiable category	unhappy	scary	Arousal	54.0911	112.8307	13.9524	5	0.0159	*	0.808	0.9397	0.9697
16	Without unclassifiable category	unhappy	scary	Dominance	-1.9028	122.985	1.3121	5	0.9337	ns.	0.808	0.9397	0.9697
16	Without unclassifiable category	unhappy	scary	Uncertainty	-3.6773	17.2623	0.5765	5	0.9891	ns.	0.808	0.9397	0.9697
16	Without unclassifiable category	unhappy	scary	MBC:Cluster2	3.2405	NA	4.8914	5	0.4293	ns.	0.808	0.9397	0.9697
16	Without unclassifiable category	unhappy	happy	Intercept	-27.8286	50.4972	NA	NA	NA	NA	0.808	0.9397	0.9697
16	Without unclassifiable category	unhappy	happy	Valence	26.6399	27.0759	3.4963	5	0.6239	ns.	0.808	0.9397	0.9697

ID	ModelType	Baseline	VECategory	TermInModel	Coefficient	StdErr	LRTest.ChiSq	DF	p	sig	McFadden	CoxSnell	Nagelkerke
16	Without unclassifiable category	unhappy	happy	Arousal	24.8905	21.7988	13.9524	5	0.0159	*	0.808	0.9397	0.9697
16	Without unclassifiable category	unhappy	happy	Dominance	-11.0324	14.6634	1.3121	5	0.9337	ns.	0.808	0.9397	0.9697
16	Without unclassifiable category	unhappy	happy	Uncertainty	-31.8657	185.5111	0.5765	5	0.9891	ns.	0.808	0.9397	0.9697
16	Without unclassifiable category	unhappy	happy	MBC:Cluster2	26.8159	34.6481	4.8914	5	0.4293	ns.	0.808	0.9397	0.9697
16	Without unclassifiable category	unhappy	confused	Intercept	-4.5855	19.2961	NA	NA	NA	NA	0.808	0.9397	0.9697
16	Without unclassifiable category	unhappy	confused	Valence	11.4082	8.2918	3.4963	5	0.6239	ns.	0.808	0.9397	0.9697
16	Without unclassifiable category	unhappy	confused	Arousal	12.2572	21.6649	13.9524	5	0.0159	*	0.808	0.9397	0.9697
16	Without unclassifiable category	unhappy	confused	Dominance	-1.4128	4.8266	1.3121	5	0.9337	ns.	0.808	0.9397	0.9697
16	Without unclassifiable category	unhappy	confused	Uncertainty	0.2686	5.2757	0.5765	5	0.9891	ns.	0.808	0.9397	0.9697
16	Without unclassifiable category	unhappy	confused	MBC:Cluster2	8.1468	31.4486	4.8914	5	0.4293	ns.	0.808	0.9397	0.9697
16	Without unclassifiable category	unhappy	uncomfortable	Intercept	-22.83	48.7931	NA	NA	NA	NA	0.808	0.9397	0.9697
16	Without unclassifiable category	unhappy	uncomfortable	Valence	0.1267	4.0607	3.4963	5	0.6239	ns.	0.808	0.9397	0.9697
16	Without unclassifiable category	unhappy	uncomfortable	Arousal	20.475	18.1913	13.9524	5	0.0159	*	0.808	0.9397	0.9697
16	Without unclassifiable category	unhappy	uncomfortable	Dominance	3.0883	5.5285	1.3121	5	0.9337	ns.	0.808	0.9397	0.9697
16	Without unclassifiable category	unhappy	uncomfortable	Uncertainty	-29.9823	199.3305	0.5765	5	0.9891	ns.	0.808	0.9397	0.9697
16	Without unclassifiable category	unhappy	uncomfortable	MBC:Cluster2	26.0902	27.1595	4.8914	5	0.4293	ns.	0.808	0.9397	0.9697
16	Without unclassifiable category	unhappy	calm	Intercept	-32.3091	374.0973	NA	NA	NA	NA	0.808	0.9397	0.9697
16	Without unclassifiable category	unhappy	calm	Valence	62.1738	715.1769	3.4963	5	0.6239	ns.	0.808	0.9397	0.9697
16	Without unclassifiable category	unhappy	calm	Arousal	-51.1406	930.7474	13.9524	5	0.0159	*	0.808	0.9397	0.9697
16	Without unclassifiable category	unhappy	calm	Dominance	-32.6072	403.0569	1.3121	5	0.9337	ns.	0.808	0.9397	0.9697
16	Without unclassifiable category	unhappy	calm	Uncertainty	-5.8174	62.2976	0.5765	5	0.9891	ns.	0.808	0.9397	0.9697
16	Without unclassifiable category	unhappy	calm	MBC:Cluster2	-32.2438	374.0973	4.8914	5	0.4293	ns.	0.808	0.9397	0.9697
17	Without unclassifiable category	Charming	Beautiful and Serene	Intercept	16.1706	22.6831	NA	NA	NA	NA	0.8563	0.9518	0.9802
17	Without unclassifiable category	Charming	Beautiful and Serene	Valence	-32.9795	25.3059	4.412	5	0.4917	ns.	0.8563	0.9518	0.9802
17	Without unclassifiable category	Charming	Beautiful and Serene	Arousal	-3.4752	6.4886	25.7028	5	0.0001	***	0.8563	0.9518	0.9802
17	Without unclassifiable category	Charming	Beautiful and Serene	Dominance	14.2586	11.344	13.1009	5	0.0225	*	0.8563	0.9518	0.9802
17	Without unclassifiable category	Charming	Beautiful and Serene	Uncertainty	-365.4841	239.8434	9.565	5	0.0885	ns.	0.8563	0.9518	0.9802
17	Without unclassifiable category	Charming	Beautiful and Serene	MBC:Cluster2	-84.3586	22.6831	0.6796	5	0.9841	ns.	0.8563	0.9518	0.9802
17	Without unclassifiable category	Charming	Creepy wonder	Intercept	-106.2052	0	NA	NA	NA	NA	0.8563	0.9518	0.9802
17	Without unclassifiable category	Charming	Creepy wonder	Valence	210.7921	0	4.412	5	0.4917	ns.	0.8563	0.9518	0.9802
17	Without unclassifiable category	Charming	Creepy wonder	Arousal	1099.8388	0	25.7028	5	0.0001	***	0.8563	0.9518	0.9802
17	Without unclassifiable category	Charming	Creepy wonder	Dominance	306.0576	0	13.1009	5	0.0225	*	0.8563	0.9518	0.9802
17	Without unclassifiable category	Charming	Creepy wonder	Uncertainty	71.4662	0	9.565	5	0.0885	ns.	0.8563	0.9518	0.9802
17	Without unclassifiable category	Charming	Creepy wonder	MBC:Cluster2	-91.9559	NA	0.6796	5	0.9841	ns.	0.8563	0.9518	0.9802
17	Without unclassifiable category	Charming	Mundane	Intercept	-111.7444	34.0532	NA	NA	NA	NA	0.8563	0.9518	0.9802
17	Without unclassifiable category	Charming	Mundane	Valence	-2.5174	11.1998	4.412	5	0.4917	ns.	0.8563	0.9518	0.9802
17	Without unclassifiable category	Charming	Mundane	Arousal	118.9356	79.6262	25.7028	5	0.0001	***	0.8563	0.9518	0.9802
17	Without unclassifiable category	Charming	Mundane	Dominance	28.6499	18.0819	13.1009	5	0.0225	*	0.8563	0.9518	0.9802
17	Without unclassifiable category	Charming	Mundane	Uncertainty	198.5809	137.8147	9.565	5	0.0885	ns.	0.8563	0.9518	0.9802
17	Without unclassifiable category	Charming	Mundane	MBC:Cluster2	214.2975	34.0274	0.6796	5	0.9841	ns.	0.8563	0.9518	0.9802

ID	ModelType	Baseline	VECategory	TermInModel	Coefficient	StdErr	LRTest.ChiSq	DF	p	sig	McFadden	CoxSnell	Nagelkerke
17	Without unclassifiable category	Charming	Sexist and Disgusting	Intercept	-43.5082	70.0211	NA	NA	NA	NA	0.8563	0.9518	0.9802
17	Without unclassifiable category	Charming	Sexist and Disgusting	Valence	-114.181	18.9128	4.412	5	0.4917	ns.	0.8563	0.9518	0.9802
17	Without unclassifiable category	Charming	Sexist and Disgusting	Arousal	417.214	94.0744	25.7028	5	0.0001	***	0.8563	0.9518	0.9802
17	Without unclassifiable category	Charming	Sexist and Disgusting	Dominance	-94.8066	10.5943	13.1009	5	0.0225	*	0.8563	0.9518	0.9802
17	Without unclassifiable category	Charming	Sexist and Disgusting	Uncertainty	-309.2166	15.9498	9.565	5	0.0885	ns.	0.8563	0.9518	0.9802
17	Without unclassifiable category	Charming	Sexist and Disgusting	MBC:Cluster2	-69.317	0	0.6796	5	0.9841	ns.	0.8563	0.9518	0.9802
17	Without unclassifiable category	Charming	Unnerving yet Boring	Intercept	259.648	39.8275	NA	NA	NA	NA	0.8563	0.9518	0.9802
17	Without unclassifiable category	Charming	Unnerving yet Boring	Valence	-81.1943	19.5903	4.412	5	0.4917	ns.	0.8563	0.9518	0.9802
17	Without unclassifiable category	Charming	Unnerving yet Boring	Arousal	439.457	75.4676	25.7028	5	0.0001	***	0.8563	0.9518	0.9802
17	Without unclassifiable category	Charming	Unnerving yet Boring	Dominance	103.8075	42.7881	13.1009	5	0.0225	*	0.8563	0.9518	0.9802
17	Without unclassifiable category	Charming	Unnerving yet Boring	Uncertainty	107.7568	9.1517	9.565	5	0.0885	ns.	0.8563	0.9518	0.9802
17	Without unclassifiable category	Charming	Unnerving yet Boring	MBC:Cluster2	-55.877	33.6927	0.6796	5	0.9841	ns.	0.8563	0.9518	0.9802
18	With unclassifiable category	Unclassifiable	most comfortable	Intercept	-50.5421	93.3325	NA	NA	NA	NA	0.375	0.7169	0.7426
18	With unclassifiable category	Unclassifiable	most comfortable	Valence	-0.3863	5.9434	4.1472	5	0.5284	ns.	0.375	0.7169	0.7426
18	With unclassifiable category	Unclassifiable	most comfortable	Arousal	-0.7334	4.89	3.1668	5	0.6743	ns.	0.375	0.7169	0.7426
18	With unclassifiable category	Unclassifiable	most comfortable	Dominance	0.7203	3.6163	3.5219	5	0.6201	ns.	0.375	0.7169	0.7426
18	With unclassifiable category	Unclassifiable	most comfortable	Uncertainty	26.1605	93.8016	4.5078	5	0.4788	ns.	0.375	0.7169	0.7426
18	With unclassifiable category	Unclassifiable	most comfortable	MBC:Cluster2	57.3044	95.1045	6.9412	5	0.2251	ns.	0.375	0.7169	0.7426
18	With unclassifiable category	Unclassifiable	most confusing	Intercept	23.8672	25.3205	NA	NA	NA	NA	0.375	0.7169	0.7426
18	With unclassifiable category	Unclassifiable	most confusing	Valence	-2.7255	6.3615	4.1472	5	0.5284	ns.	0.375	0.7169	0.7426
18	With unclassifiable category	Unclassifiable	most confusing	Arousal	-3.9966	5.3279	3.1668	5	0.6743	ns.	0.375	0.7169	0.7426
18	With unclassifiable category	Unclassifiable	most confusing	Dominance	0.8447	3.7119	3.5219	5	0.6201	ns.	0.375	0.7169	0.7426
18	With unclassifiable category	Unclassifiable	most confusing	Uncertainty	10.3519	94.3581	4.5078	5	0.4788	ns.	0.375	0.7169	0.7426
18	With unclassifiable category	Unclassifiable	most confusing	MBC:Cluster2	-21.9753	27.5336	6.9412	5	0.2251	ns.	0.375	0.7169	0.7426
18	With unclassifiable category	Unclassifiable	most realistic	Intercept	27.8302	25.9068	NA	NA	NA	NA	0.375	0.7169	0.7426
18	With unclassifiable category	Unclassifiable	most realistic	Valence	6.6411	7.3765	4.1472	5	0.5284	ns.	0.375	0.7169	0.7426
18	With unclassifiable category	Unclassifiable	most realistic	Arousal	-2.9511	5.3858	3.1668	5	0.6743	ns.	0.375	0.7169	0.7426
18	With unclassifiable category	Unclassifiable	most realistic	Dominance	-5.0285	4.9187	3.5219	5	0.6201	ns.	0.375	0.7169	0.7426

ID	ModelType	Baseline	VECategory	TermInModel	Coefficient	StdErr	LRTest.ChiSq	DF	p	sig	McFadden	CoxSnell	Nagelkerke
18	With unclassifiable category	Unclassifiable	most realistic	Uncertainty	10.817	94.4068	4.5078	5	0.4788	ns.	0.375	0.7169	0.7426
18	With unclassifiable category	Unclassifiable	most realistic	MBC:Cluster2	-29.1283	28.7557	6.9412	5	0.2251	ns.	0.375	0.7169	0.7426
18	With unclassifiable category	Unclassifiable	most uncomfortable	Intercept	19.6707	86.6542	NA	NA	NA	NA	0.375	0.7169	0.7426
18	With unclassifiable category	Unclassifiable	most uncomfortable	Valence	2.5417	6.7573	4.1472	5	0.5284	ns.	0.375	0.7169	0.7426
18	With unclassifiable category	Unclassifiable	most uncomfortable	Arousal	-0.724	5.3955	3.1668	5	0.6743	ns.	0.375	0.7169	0.7426
18	With unclassifiable category	Unclassifiable	most uncomfortable	Dominance	0.3515	4.7155	3.5219	5	0.6201	ns.	0.375	0.7169	0.7426
18	With unclassifiable category	Unclassifiable	most uncomfortable	Uncertainty	-30.5184	412.2363	4.5078	5	0.4788	ns.	0.375	0.7169	0.7426
18	With unclassifiable category	Unclassifiable	most uncomfortable	MBC:Cluster2	-29.8616	28.2025	6.9412	5	0.2251	ns.	0.375	0.7169	0.7426
18	With unclassifiable category	Unclassifiable	ones in which I found a present?	Intercept	23.7089	26.0066	NA	NA	NA	NA	0.375	0.7169	0.7426
18	With unclassifiable category	Unclassifiable	ones in which I found a present?	Valence	0.347	6.9838	4.1472	5	0.5284	ns.	0.375	0.7169	0.7426
18	With unclassifiable category	Unclassifiable	ones in which I found a present?	Arousal	-4.5197	5.6664	3.1668	5	0.6743	ns.	0.375	0.7169	0.7426
18	With unclassifiable category	Unclassifiable	ones in which I found a present?	Dominance	-2.3041	4.6025	3.5219	5	0.6201	ns.	0.375	0.7169	0.7426
18	With unclassifiable category	Unclassifiable	ones in which I found a present?	Uncertainty	-4.4878	95.4601	4.5078	5	0.4788	ns.	0.375	0.7169	0.7426
18	With unclassifiable category	Unclassifiable	ones in which I found a present?	MBC:Cluster2	-27.0422	28.3458	6.9412	5	0.2251	ns.	0.375	0.7169	0.7426
19	Without unclassifiable category	calm	Active	Intercept	19.5885	14.6592	NA	NA	NA	NA	0.5187	0.7973	0.8358
19	Without unclassifiable category	calm	Active	Valence	2.9805	3.9923	4.6575	4	0.3243	ns.	0.5187	0.7973	0.8358
19	Without unclassifiable category	calm	Active	Arousal	2.3536	4.089	13.5651	4	0.0088	**	0.5187	0.7973	0.8358
19	Without unclassifiable category	calm	Active	Dominance	-1.2191	2.6458	3.3395	4	0.5027	ns.	0.5187	0.7973	0.8358
19	Without unclassifiable category	calm	Active	Uncertainty	7.8481	25.068	3.0301	4	0.5528	ns.	0.5187	0.7973	0.8358
19	Without unclassifiable category	calm	Active	MBC:Cluster2	-18.7404	14.4956	7.759	4	0.1008	ns.	0.5187	0.7973	0.8358
19	Without unclassifiable category	calm	Horrified	Intercept	-23.3932	33.7814	NA	NA	NA	NA	0.5187	0.7973	0.8358
19	Without unclassifiable category	calm	Horrified	Valence	-4.3854	7.443	4.6575	4	0.3243	ns.	0.5187	0.7973	0.8358
19	Without unclassifiable category	calm	Horrified	Arousal	-22.7859	54.3555	13.5651	4	0.0088	**	0.5187	0.7973	0.8358
19	Without unclassifiable category	calm	Horrified	Dominance	-52.8758	67.5294	3.3395	4	0.5027	ns.	0.5187	0.7973	0.8358
19	Without unclassifiable category	calm	Horrified	Uncertainty	-28.0549	88.5263	3.0301	4	0.5528	ns.	0.5187	0.7973	0.8358
19	Without unclassifiable category	calm	Horrified	MBC:Cluster2	-48.0882	0	7.759	4	0.1008	ns.	0.5187	0.7973	0.8358
19	Without unclassifiable category	calm	Magical / surreal	Intercept	3.5012	15.5135	NA	NA	NA	NA	0.5187	0.7973	0.8358
19	Without unclassifiable category	calm	Magical / surreal	Valence	5.3396	4.3918	4.6575	4	0.3243	ns.	0.5187	0.7973	0.8358
19	Without unclassifiable category	calm	Magical / surreal	Arousal	19.1374	12.719	13.5651	4	0.0088	**	0.5187	0.7973	0.8358
19	Without unclassifiable category	calm	Magical / surreal	Dominance	-2.3388	3.1665	3.3395	4	0.5027	ns.	0.5187	0.7973	0.8358

ID	ModelType	Baseline	VECategory	TermInModel	Coefficient	StdErr	LRTest.ChiSq	DF	p	sig	McFadden	CoxSnell	Nagelkerke
19	Without unclassifiable category	calm	Magical / surreal	Uncertainty	1.1096	76.3037	3.0301	4	0.5528	ns.	0.5187	0.7973	0.8358
19	Without unclassifiable category	calm	Magical / surreal	MBC:Cluster2	5.5738	10.4242	7.759	4	0.1008	ns.	0.5187	0.7973	0.8358
19	Without unclassifiable category	calm	Unsettled	Intercept	-17.7901	22.0489	NA	NA	NA	NA	0.5187	0.7973	0.8358
19	Without unclassifiable category	calm	Unsettled	Valence	1.3107	3.4268	4.6575	4	0.3243	ns.	0.5187	0.7973	0.8358
19	Without unclassifiable category	calm	Unsettled	Arousal	18.8396	12.2954	13.5651	4	0.0088	**	0.5187	0.7973	0.8358
19	Without unclassifiable category	calm	Unsettled	Dominance	-0.9264	2.5299	3.3395	4	0.5027	ns.	0.5187	0.7973	0.8358
19	Without unclassifiable category	calm	Unsettled	Uncertainty	-85.4862	109.774	3.0301	4	0.5528	ns.	0.5187	0.7973	0.8358
19	Without unclassifiable category	calm	Unsettled	MBC:Cluster2	10.0067	10.0858	7.759	4	0.1008	ns.	0.5187	0.7973	0.8358
20	Without unclassifiable category	Comfortable	scary	Intercept	-53.9629	2.5364	NA	NA	NA	NA	0.9199	0.9445	0.9871
20	Without unclassifiable category	Comfortable	scary	Valence	-57.1538	1.5909	0.4797	4	0.9754	ns.	0.9199	0.9445	0.9871
20	Without unclassifiable category	Comfortable	scary	Arousal	122.4446	5.2557	2.3433	4	0.6729	ns.	0.9199	0.9445	0.9871
20	Without unclassifiable category	Comfortable	scary	Dominance	2.8422	6.7488	6.1816	4	0.186	ns.	0.9199	0.9445	0.9871
20	Without unclassifiable category	Comfortable	scary	Uncertainty	-76.439	0.5824	0.6351	4	0.9591	ns.	0.9199	0.9445	0.9871
20	Without unclassifiable category	Comfortable	scary	MBC:Cluster2	51.5369	0	0.0012	4	1	ns.	0.9199	0.9445	0.9871
20	Without unclassifiable category	Comfortable	peaceful	Intercept	-131.872	22.162	NA	NA	NA	NA	0.9199	0.9445	0.9871
20	Without unclassifiable category	Comfortable	peaceful	Valence	168.2363	19.8284	0.4797	4	0.9754	ns.	0.9199	0.9445	0.9871
20	Without unclassifiable category	Comfortable	peaceful	Arousal	-57.4598	15.2204	2.3433	4	0.6729	ns.	0.9199	0.9445	0.9871
20	Without unclassifiable category	Comfortable	peaceful	Dominance	200.6554	19.7739	6.1816	4	0.186	ns.	0.9199	0.9445	0.9871
20	Without unclassifiable category	Comfortable	peaceful	Uncertainty	1.4807	5.0721	0.6351	4	0.9591	ns.	0.9199	0.9445	0.9871
20	Without unclassifiable category	Comfortable	peaceful	MBC:Cluster2	-169.1507	22.162	0.0012	4	1	ns.	0.9199	0.9445	0.9871
20	Without unclassifiable category	Comfortable	interesting / curious	Intercept	132.9506	22.162	NA	NA	NA	NA	0.9199	0.9445	0.9871
20	Without unclassifiable category	Comfortable	interesting / curious	Valence	-79.4791	19.8284	0.4797	4	0.9754	ns.	0.9199	0.9445	0.9871
20	Without unclassifiable category	Comfortable	interesting / curious	Arousal	50.5086	15.2204	2.3433	4	0.6729	ns.	0.9199	0.9445	0.9871
20	Without unclassifiable category	Comfortable	interesting / curious	Dominance	218.0913	19.7739	6.1816	4	0.186	ns.	0.9199	0.9445	0.9871
20	Without unclassifiable category	Comfortable	interesting / curious	Uncertainty	-44.1419	5.0721	0.6351	4	0.9591	ns.	0.9199	0.9445	0.9871
20	Without unclassifiable category	Comfortable	interesting / curious	MBC:Cluster2	-174.6013	22.162	0.0012	4	1	ns.	0.9199	0.9445	0.9871
20	Without unclassifiable category	Comfortable	Thrilling	Intercept	-67.5207	2.5364	NA	NA	NA	NA	0.9199	0.9445	0.9871
20	Without unclassifiable category	Comfortable	Thrilling	Valence	-54.9714	1.5909	0.4797	4	0.9754	ns.	0.9199	0.9445	0.9871
20	Without unclassifiable category	Comfortable	Thrilling	Arousal	110.6411	5.2557	2.3433	4	0.6729	ns.	0.9199	0.9445	0.9871
20	Without unclassifiable category	Comfortable	Thrilling	Dominance	-11.0882	6.7488	6.1816	4	0.186	ns.	0.9199	0.9445	0.9871
20	Without unclassifiable category	Comfortable	Thrilling	Uncertainty	-138.4423	0.5824	0.6351	4	0.9591	ns.	0.9199	0.9445	0.9871
20	Without unclassifiable category	Comfortable	Thrilling	MBC:Cluster2	-15.3199	0	0.0012	4	1	ns.	0.9199	0.9445	0.9871
21	Without unclassifiable category	Happy and Excited	bored	Intercept	-1503.3475	12290.7134	NA	NA	NA	NA	1	0.977	1
21	Without unclassifiable category	Happy and Excited	bored	Valence	-457.65	15874.7298	17.2952	6	0.0083	**	1	0.977	1
21	Without unclassifiable category	Happy and Excited	bored	Arousal	-3767.5671	31913.7855	17.5269	6	0.0075	**	1	0.977	1
21	Without unclassifiable category	Happy and Excited	bored	Dominance	-2227.5359	21098.4029	9.1829	6	0.1636	ns.	1	0.977	1
21	Without unclassifiable category	Happy and Excited	bored	Uncertainty	-398.8183	16914.5152	6.0607	6	0.4164	ns.	1	0.977	1
21	Without unclassifiable category	Happy and Excited	bored	MBC:Cluster2	74.8847	12290.7134	0	6	1	ns.	1	0.977	1

ID	ModelType	Baseline	VECategory	TermInModel	Coefficient	StdErr	LRTest.ChiSq	DF	p	sig	McFadden	CoxSnell	Nagelkerke
21	Without unclassifiable category	Happy and Excited	scary	Intercept	-2896.8111	0	NA	NA	NA	NA	1	0.977	1
21	Without unclassifiable category	Happy and Excited	scary	Valence	-2931.2425	0	17.2952	6	0.0083	**	1	0.977	1
21	Without unclassifiable category	Happy and Excited	scary	Arousal	419.8073	0	17.5269	6	0.0075	**	1	0.977	1
21	Without unclassifiable category	Happy and Excited	scary	Dominance	1057.2379	0	9.1829	6	0.1636	ns.	1	0.977	1
21	Without unclassifiable category	Happy and Excited	scary	Uncertainty	595.0605	0	6.0607	6	0.4164	ns.	1	0.977	1
21	Without unclassifiable category	Happy and Excited	scary	MBC:Cluster2	653.7813	0	0	6	1	ns.	1	0.977	1
21	Without unclassifiable category	Happy and Excited	relaxed	Intercept	-1453.8188	2107.3902	NA	NA	NA	NA	1	0.977	1
21	Without unclassifiable category	Happy and Excited	relaxed	Valence	3121.6864	524.7304	17.2952	6	0.0083	**	1	0.977	1
21	Without unclassifiable category	Happy and Excited	relaxed	Arousal	-1350.5911	85.1366	17.5269	6	0.0075	**	1	0.977	1
21	Without unclassifiable category	Happy and Excited	relaxed	Dominance	529.6455	11.1262	9.1829	6	0.1636	ns.	1	0.977	1
21	Without unclassifiable category	Happy and Excited	relaxed	Uncertainty	1291.8945	9863.1247	6.0607	6	0.4164	ns.	1	0.977	1
21	Without unclassifiable category	Happy and Excited	relaxed	MBC:Cluster2	-2212.3638	0	0	6	1	ns.	1	0.977	1
21	Without unclassifiable category	Happy and Excited	confused	Intercept	658.0997	656.2366	NA	NA	NA	NA	1	0.977	1
21	Without unclassifiable category	Happy and Excited	confused	Valence	-190.9457	6329.6709	17.2952	6	0.0083	**	1	0.977	1
21	Without unclassifiable category	Happy and Excited	confused	Arousal	-1311.5738	4028.9505	17.5269	6	0.0075	**	1	0.977	1
21	Without unclassifiable category	Happy and Excited	confused	Dominance	-1137.3424	2209.0641	9.1829	6	0.1636	ns.	1	0.977	1
21	Without unclassifiable category	Happy and Excited	confused	Uncertainty	243.1728	148.4859	6.0607	6	0.4164	ns.	1	0.977	1
21	Without unclassifiable category	Happy and Excited	confused	MBC:Cluster2	-1141.7149	0	0	6	1	ns.	1	0.977	1
21	Without unclassifiable category	Happy and Excited	Respectful	Intercept	-2216.8256	10184.0662	NA	NA	NA	NA	1	0.977	1
21	Without unclassifiable category	Happy and Excited	Respectful	Valence	93.3693	15350.1132	17.2952	6	0.0083	**	1	0.977	1
21	Without unclassifiable category	Happy and Excited	Respectful	Arousal	-3602.8687	31828.7006	17.5269	6	0.0075	**	1	0.977	1
21	Without unclassifiable category	Happy and Excited	Respectful	Dominance	-2475.3171	21087.2795	9.1829	6	0.1636	ns.	1	0.977	1
21	Without unclassifiable category	Happy and Excited	Respectful	Uncertainty	1313.789	7052.3491	6.0607	6	0.4164	ns.	1	0.977	1
21	Without unclassifiable category	Happy and Excited	Respectful	MBC:Cluster2	1069.1945	12290.7134	0	6	1	ns.	1	0.977	1
21	Without unclassifiable category	Happy and Excited	Sexy Put-off	Intercept	-1020.1959	656.2366	NA	NA	NA	NA	1	0.977	1
21	Without unclassifiable category	Happy and Excited	Sexy Put-off	Valence	-462.3506	6329.6709	17.2952	6	0.0083	**	1	0.977	1

ID	ModelType	Baseline	VECategory	TermInModel	Coefficient	StdErr	LRTest.ChiSq	DF	p	sig	McFadden	CoxSnell	Nagelkerke
21	Without unclassifiable category	Happy and Excited	Sexy Put-off	Arousal	-1846.6394	4028.9505	17.5269	6	0.0075	**	1	0.977	1
21	Without unclassifiable category	Happy and Excited	Sexy Put-off	Dominance	-2818.7822	2209.0641	9.1829	6	0.1636	ns.	1	0.977	1
21	Without unclassifiable category	Happy and Excited	Sexy Put-off	Uncertainty	-835.7864	148.4859	6.0607	6	0.4164	ns.	1	0.977	1
21	Without unclassifiable category	Happy and Excited	Sexy Put-off	MBC:Cluster2	-538.7343	0	0	6	1	ns.	1	0.977	1
22	With unclassifiable category	Unclassifiable	sex	Intercept	-172.9249	220.1029	NA	NA	NA	NA	0.8553	0.9566	0.9817
22	With unclassifiable category	Unclassifiable	sex	Valence	-21.3292	32.308	21.5097	6	0.0015	**	0.8553	0.9566	0.9817
22	With unclassifiable category	Unclassifiable	sex	Arousal	-64.2908	88.8803	25.0794	6	0.0003	***	0.8553	0.9566	0.9817
22	With unclassifiable category	Unclassifiable	sex	Dominance	-220.2023	216.8784	35.2959	6	0	***	0.8553	0.9566	0.9817
22	With unclassifiable category	Unclassifiable	sex	Uncertainty	-133.5566	49.6557	13.1535	6	0.0407	*	0.8553	0.9566	0.9817
22	With unclassifiable category	Unclassifiable	sex	MBC:Cluster2	-205.3431	0	23.7672	6	0.0006	***	0.8553	0.9566	0.9817
22	With unclassifiable category	Unclassifiable	Childlike	Intercept	-193.6925	36.6364	NA	NA	NA	NA	0.8553	0.9566	0.9817
22	With unclassifiable category	Unclassifiable	Childlike	Valence	7.235	13.7582	21.5097	6	0.0015	**	0.8553	0.9566	0.9817
22	With unclassifiable category	Unclassifiable	Childlike	Arousal	13.4039	35.6958	25.0794	6	0.0003	***	0.8553	0.9566	0.9817
22	With unclassifiable category	Unclassifiable	Childlike	Dominance	-0.4177	5.9172	35.2959	6	0	***	0.8553	0.9566	0.9817
22	With unclassifiable category	Unclassifiable	Childlike	Uncertainty	14.4007	344.4386	13.1535	6	0.0407	*	0.8553	0.9566	0.9817
22	With unclassifiable category	Unclassifiable	Childlike	MBC:Cluster2	199.933	36.6364	23.7672	6	0.0006	***	0.8553	0.9566	0.9817
22	With unclassifiable category	Unclassifiable	Desolate	Intercept	97.0651	179.3835	NA	NA	NA	NA	0.8553	0.9566	0.9817
22	With unclassifiable category	Unclassifiable	Desolate	Valence	10.519	13.9925	21.5097	6	0.0015	**	0.8553	0.9566	0.9817
22	With unclassifiable category	Unclassifiable	Desolate	Arousal	6.0652	36.6817	25.0794	6	0.0003	***	0.8553	0.9566	0.9817
22	With unclassifiable category	Unclassifiable	Desolate	Dominance	1.5886	6.457	35.2959	6	0	***	0.8553	0.9566	0.9817
22	With unclassifiable category	Unclassifiable	Desolate	Uncertainty	240.7911	276.0303	13.1535	6	0.0407	*	0.8553	0.9566	0.9817
22	With unclassifiable category	Unclassifiable	Desolate	MBC:Cluster2	-48.5339	186.1147	23.7672	6	0.0006	***	0.8553	0.9566	0.9817
22	With unclassifiable category	Unclassifiable	fantasylike	Intercept	-81.5176	112.1751	NA	NA	NA	NA	0.8553	0.9566	0.9817
22	With unclassifiable category	Unclassifiable	fantasylike	Valence	-3.4709	51.6673	21.5097	6	0.0015	**	0.8553	0.9566	0.9817
22	With unclassifiable category	Unclassifiable	fantasylike	Arousal	-195.0051	260.8166	25.0794	6	0.0003	***	0.8553	0.9566	0.9817
22	With unclassifiable category	Unclassifiable	fantasylike	Dominance	-37.5974	51.4292	35.2959	6	0	***	0.8553	0.9566	0.9817
22	With unclassifiable category	Unclassifiable	fantasylike	Uncertainty	-152.7953	532.424	13.1535	6	0.0407	*	0.8553	0.9566	0.9817
22	With unclassifiable category	Unclassifiable	fantasylike	MBC:Cluster2	-51.3685	112.1751	23.7672	6	0.0006	***	0.8553	0.9566	0.9817
22	With unclassifiable category	Unclassifiable	Iconoclasm	Intercept	-353.6848	65.6456	NA	NA	NA	NA	0.8553	0.9566	0.9817
22	With unclassifiable category	Unclassifiable	Iconoclasm	Valence	39.662	84.975	21.5097	6	0.0015	**	0.8553	0.9566	0.9817
22	With unclassifiable category	Unclassifiable	Iconoclasm	Arousal	-386.8742	252.9553	25.0794	6	0.0003	***	0.8553	0.9566	0.9817
22	With unclassifiable category	Unclassifiable	Iconoclasm	Dominance	-769.3959	98.7606	35.2959	6	0	***	0.8553	0.9566	0.9817
22	With unclassifiable category	Unclassifiable	Iconoclasm	Uncertainty	114.178	97.3586	13.1535	6	0.0407	*	0.8553	0.9566	0.9817
22	With unclassifiable category	Unclassifiable	Iconoclasm	MBC:Cluster2	40.1028	0	23.7672	6	0.0006	***	0.8553	0.9566	0.9817
22	With unclassifiable category	Unclassifiable	Italian influence	Intercept	321.6085	125.8437	NA	NA	NA	NA	0.8553	0.9566	0.9817
22	With unclassifiable category	Unclassifiable	Italian influence	Valence	312.8045	389.2742	21.5097	6	0.0015	**	0.8553	0.9566	0.9817
22	With unclassifiable category	Unclassifiable	Italian influence	Arousal	-190.5874	267.3664	25.0794	6	0.0003	***	0.8553	0.9566	0.9817
22	With unclassifiable category	Unclassifiable	Italian influence	Dominance	-162.8075	207.4933	35.2959	6	0	***	0.8553	0.9566	0.9817
22	With unclassifiable category	Unclassifiable	Italian influence	Uncertainty	211.7578	274.2382	13.1535	6	0.0407	*	0.8553	0.9566	0.9817
22	With unclassifiable category	Unclassifiable	Italian influence	MBC:Cluster2	-538.8025	464.1342	23.7672	6	0.0006	***	0.8553	0.9566	0.9817
23	Without unclassifiable category	Annoyed	scary	Intercept	-26.6031	4.9098	NA	NA	NA	NA	0.9325	0.9804	0.9951
23	Without unclassifiable category	Annoyed	scary	Valence	0.4101	2.825	3.2266	8	0.9193	ns.	0.9325	0.9804	0.9951

ID	ModelType	Baseline	VECategory	TermInModel	Coefficient	StdErr	LRTest.ChiSq	DF	p	sig	McFadden	CoxSnell	Nagelkerke
23	Without unclassifiable category	Annoyed	scary	Arousal	5.2197	7.2332	13.5018	8	0.0957	ns.	0.9325	0.9804	0.9951
23	Without unclassifiable category	Annoyed	scary	Dominance	2.8639	9.7161	8.552	8	0.3815	ns.	0.9325	0.9804	0.9951
23	Without unclassifiable category	Annoyed	scary	Uncertainty	-103.1986	1.126	9.2578	8	0.321	ns.	0.9325	0.9804	0.9951
23	Without unclassifiable category	Annoyed	scary	MBC:Cluster2	1086.6971	0	0.0402	8	1	ns.	0.9325	0.9804	0.9951
23	Without unclassifiable category	Annoyed	happy	Intercept	-3559.9549	0	NA	NA	NA	NA	0.9325	0.9804	0.9951
23	Without unclassifiable category	Annoyed	happy	Valence	7508.9176	0	3.2266	8	0.9193	ns.	0.9325	0.9804	0.9951
23	Without unclassifiable category	Annoyed	happy	Arousal	-714.8041	0	13.5018	8	0.0957	ns.	0.9325	0.9804	0.9951
23	Without unclassifiable category	Annoyed	happy	Dominance	6333.5076	0	8.552	8	0.3815	ns.	0.9325	0.9804	0.9951
23	Without unclassifiable category	Annoyed	happy	Uncertainty	1206.9209	0	9.2578	8	0.321	ns.	0.9325	0.9804	0.9951
23	Without unclassifiable category	Annoyed	happy	MBC:Cluster2	-4265.2789	0	0.0402	8	1	ns.	0.9325	0.9804	0.9951
23	Without unclassifiable category	Annoyed	curious	Intercept	2157.8982	77.4771	NA	NA	NA	NA	0.9325	0.9804	0.9951
23	Without unclassifiable category	Annoyed	curious	Valence	1118.0806	60.4588	3.2266	8	0.9193	ns.	0.9325	0.9804	0.9951
23	Without unclassifiable category	Annoyed	curious	Arousal	758.8266	50.5272	13.5018	8	0.0957	ns.	0.9325	0.9804	0.9951
23	Without unclassifiable category	Annoyed	curious	Dominance	2706.0893	59.3122	8.552	8	0.3815	ns.	0.9325	0.9804	0.9951
23	Without unclassifiable category	Annoyed	curious	Uncertainty	-648.2582	17.6591	9.2578	8	0.321	ns.	0.9325	0.9804	0.9951
23	Without unclassifiable category	Annoyed	curious	MBC:Cluster2	1403.8939	77.4771	0.0402	8	1	ns.	0.9325	0.9804	0.9951
23	Without unclassifiable category	Annoyed	Adventurous	Intercept	-359.8194	483.5922	NA	NA	NA	NA	0.9325	0.9804	0.9951
23	Without unclassifiable category	Annoyed	Adventurous	Valence	330.9127	647.128	3.2266	8	0.9193	ns.	0.9325	0.9804	0.9951
23	Without unclassifiable category	Annoyed	Adventurous	Arousal	-3823.4187	1841.805	13.5018	8	0.0957	ns.	0.9325	0.9804	0.9951
23	Without unclassifiable category	Annoyed	Adventurous	Dominance	2486.1837	521.1	8.552	8	0.3815	ns.	0.9325	0.9804	0.9951
23	Without unclassifiable category	Annoyed	Adventurous	Uncertainty	-831.4608	166.2622	9.2578	8	0.321	ns.	0.9325	0.9804	0.9951
23	Without unclassifiable category	Annoyed	Adventurous	MBC:Cluster2	1686.2258	483.5922	0.0402	8	1	ns.	0.9325	0.9804	0.9951
23	Without unclassifiable category	Annoyed	Energetic	Intercept	1179.006	NA	NA	NA	NA	NA	0.9325	0.9804	0.9951
23	Without unclassifiable category	Annoyed	Energetic	Valence	-1507.3347	0	3.2266	8	0.9193	ns.	0.9325	0.9804	0.9951
23	Without unclassifiable category	Annoyed	Energetic	Arousal	-5231.275	0	13.5018	8	0.0957	ns.	0.9325	0.9804	0.9951
23	Without unclassifiable category	Annoyed	Energetic	Dominance	-844.5524	0	8.552	8	0.3815	ns.	0.9325	0.9804	0.9951
23	Without unclassifiable category	Annoyed	Energetic	Uncertainty	1197.3973	0	9.2578	8	0.321	ns.	0.9325	0.9804	0.9951
23	Without unclassifiable category	Annoyed	Energetic	MBC:Cluster2	-1873.7934	0	0.0402	8	1	ns.	0.9325	0.9804	0.9951
23	Without unclassifiable category	Annoyed	Inspired	Intercept	-1239.0748	484.4099	NA	NA	NA	NA	0.9325	0.9804	0.9951
23	Without unclassifiable category	Annoyed	Inspired	Valence	1090.45	645.2118	3.2266	8	0.9193	ns.	0.9325	0.9804	0.9951
23	Without unclassifiable category	Annoyed	Inspired	Arousal	-3716.5612	1841.6005	13.5018	8	0.0957	ns.	0.9325	0.9804	0.9951
23	Without unclassifiable category	Annoyed	Inspired	Dominance	2082.3584	523.4818	8.552	8	0.3815	ns.	0.9325	0.9804	0.9951
23	Without unclassifiable category	Annoyed	Inspired	Uncertainty	1769.5649	165.8395	9.2578	8	0.321	ns.	0.9325	0.9804	0.9951
23	Without unclassifiable category	Annoyed	Inspired	MBC:Cluster2	2935.5076	484.4099	0.0402	8	1	ns.	0.9325	0.9804	0.9951
23	Without unclassifiable category	Annoyed	Intimidated	Intercept	-2186.4584	0	NA	NA	NA	NA	0.9325	0.9804	0.9951
23	Without unclassifiable category	Annoyed	Intimidated	Valence	3606.0542	0	3.2266	8	0.9193	ns.	0.9325	0.9804	0.9951
23	Without unclassifiable category	Annoyed	Intimidated	Arousal	-6133.4621	0	13.5018	8	0.0957	ns.	0.9325	0.9804	0.9951
23	Without unclassifiable category	Annoyed	Intimidated	Dominance	-5287.4371	0	8.552	8	0.3815	ns.	0.9325	0.9804	0.9951
23	Without unclassifiable category	Annoyed	Intimidated	Uncertainty	-108.6159	0	9.2578	8	0.321	ns.	0.9325	0.9804	0.9951
23	Without unclassifiable category	Annoyed	Intimidated	MBC:Cluster2	47.7856	0	0.0402	8	1	ns.	0.9325	0.9804	0.9951
23	Without unclassifiable category	Annoyed	Playful	Intercept	-1162.0961	0	NA	NA	NA	NA	0.9325	0.9804	0.9951
23	Without unclassifiable category	Annoyed	Playful	Valence	3874.3615	0	3.2266	8	0.9193	ns.	0.9325	0.9804	0.9951
23	Without unclassifiable category	Annoyed	Playful	Arousal	5119.7436	0	13.5018	8	0.0957	ns.	0.9325	0.9804	0.9951
23	Without unclassifiable category	Annoyed	Playful	Dominance	5230.4219	0	8.552	8	0.3815	ns.	0.9325	0.9804	0.9951
23	Without unclassifiable category	Annoyed	Playful	Uncertainty	610.9126	0	9.2578	8	0.321	ns.	0.9325	0.9804	0.9951
23	Without unclassifiable category	Annoyed	Playful	MBC:Cluster2	1451.754	0	0.0402	8	1	ns.	0.9325	0.9804	0.9951
24	Without unclassifiable category	pleasant	scary	Intercept	-205.7507	557.5444	NA	NA	NA	NA	0.931	0.957	0.9907
24	Without unclassifiable category	pleasant	scary	Valence	68.8186	1464.175	1.6029	5	0.9009	ns.	0.931	0.957	0.9907
24	Without unclassifiable category	pleasant	scary	Arousal	1596.8406	711.513	37.0031	5	0	***	0.931	0.957	0.9907
24	Without unclassifiable category	pleasant	scary	Dominance	859.2865	296.7303	22.0454	5	0.0005	***	0.931	0.957	0.9907
24	Without unclassifiable category	pleasant	scary	Uncertainty	104.208	126.8036	-0.8253	5	1	ns.	0.931	0.957	0.9907
24	Without unclassifiable category	pleasant	scary	MBC:Cluster2	-202.6213	0	5.3123	5	0.379	ns.	0.931	0.957	0.9907

ID	ModelType	Baseline	VECategory	TermInModel	Coefficient	StdErr	LRTest.ChiSq	DF	p	sig	McFadden	CoxSnell	Nagelkerke
24	Without unclassifiable category	pleasant	sick	Intercept	-57.2334	1972.0225	NA	NA	NA	NA	0.931	0.957	0.9907
24	Without unclassifiable category	pleasant	sick	Valence	73.681	738.8731	1.6029	5	0.9009	ns.	0.931	0.957	0.9907
24	Without unclassifiable category	pleasant	sick	Arousal	643.5989	960.7657	37.0031	5	0	***	0.931	0.957	0.9907
24	Without unclassifiable category	pleasant	sick	Dominance	-121.832	3695.0445	22.0454	5	0.0005	***	0.931	0.957	0.9907
24	Without unclassifiable category	pleasant	sick	Uncertainty	-356.2113	447.9236	-0.8253	5	1	ns.	0.931	0.957	0.9907
24	Without unclassifiable category	pleasant	sick	MBC:Cluster2	-250.1059	0	5.3123	5	0.379	ns.	0.931	0.957	0.9907
24	Without unclassifiable category	pleasant	claustrophobic	Intercept	148.7856	396.6364	NA	NA	NA	NA	0.931	0.957	0.9907
24	Without unclassifiable category	pleasant	claustrophobic	Valence	102.9435	547.7488	1.6029	5	0.9009	ns.	0.931	0.957	0.9907
24	Without unclassifiable category	pleasant	claustrophobic	Arousal	667.0597	898.059	37.0031	5	0	***	0.931	0.957	0.9907
24	Without unclassifiable category	pleasant	claustrophobic	Dominance	29.1848	488.7131	22.0454	5	0.0005	***	0.931	0.957	0.9907
24	Without unclassifiable category	pleasant	claustrophobic	Uncertainty	-137.2748	111.8275	-0.8253	5	1	ns.	0.931	0.957	0.9907
24	Without unclassifiable category	pleasant	claustrophobic	MBC:Cluster2	142.8633	188.7783	5.3123	5	0.379	ns.	0.931	0.957	0.9907
24	Without unclassifiable category	pleasant	lifeless	Intercept	186.1158	1070.7747	NA	NA	NA	NA	0.931	0.957	0.9907
24	Without unclassifiable category	pleasant	lifeless	Valence	110.2308	902.3116	1.6029	5	0.9009	ns.	0.931	0.957	0.9907
24	Without unclassifiable category	pleasant	lifeless	Arousal	943.5096	464.8291	37.0031	5	0	***	0.931	0.957	0.9907
24	Without unclassifiable category	pleasant	lifeless	Dominance	340.8896	1184.987	22.0454	5	0.0005	***	0.931	0.957	0.9907
24	Without unclassifiable category	pleasant	lifeless	Uncertainty	-45.9127	247.1684	-0.8253	5	1	ns.	0.931	0.957	0.9907
24	Without unclassifiable category	pleasant	lifeless	MBC:Cluster2	-161.9441	84.8871	5.3123	5	0.379	ns.	0.931	0.957	0.9907
24	Without unclassifiable category	pleasant	lively	Intercept	382.1482	98.0502	NA	NA	NA	NA	0.931	0.957	0.9907
24	Without unclassifiable category	pleasant	lively	Valence	97.2005	362.9669	1.6029	5	0.9009	ns.	0.931	0.957	0.9907
24	Without unclassifiable category	pleasant	lively	Arousal	157.2354	689.8072	37.0031	5	0	***	0.931	0.957	0.9907
24	Without unclassifiable category	pleasant	lively	Dominance	-30.0172	158.2299	22.0454	5	0.0005	***	0.931	0.957	0.9907
24	Without unclassifiable category	pleasant	lively	Uncertainty	174.261	800.2889	-0.8253	5	1	ns.	0.931	0.957	0.9907
24	Without unclassifiable category	pleasant	lively	MBC:Cluster2	-293.9291	98.0502	5.3123	5	0.379	ns.	0.931	0.957	0.9907
25	With unclassifiable category	Unclassifiable	scary	Intercept	-135.2352	32.7862	NA	NA	NA	NA	0.8322	0.9267	0.9686
25	With unclassifiable category	Unclassifiable	scary	Valence	84.4614	59.6205	10.3277	5	0.0665	ns.	0.8322	0.9267	0.9686
25	With unclassifiable category	Unclassifiable	scary	Arousal	134.5942	95.2372	13.6047	5	0.0183	*	0.8322	0.9267	0.9686
25	With unclassifiable category	Unclassifiable	scary	Dominance	-82.7005	51.8692	3.0143	5	0.6978	ns.	0.8322	0.9267	0.9686
25	With unclassifiable category	Unclassifiable	scary	Uncertainty	41.1246	6.6561	2.2091	5	0.8195	ns.	0.8322	0.9267	0.9686
25	With unclassifiable category	Unclassifiable	scary	MBC:Cluster2	266.9286	55.9312	15.1802	5	0.0096	**	0.8322	0.9267	0.9686
25	With unclassifiable category	Unclassifiable	Worried	Intercept	30.2045	137.2795	NA	NA	NA	NA	0.8322	0.9267	0.9686
25	With unclassifiable category	Unclassifiable	Worried	Valence	53.6138	112.8829	10.3277	5	0.0665	ns.	0.8322	0.9267	0.9686
25	With unclassifiable category	Unclassifiable	Worried	Arousal	104.1475	74.721	13.6047	5	0.0183	*	0.8322	0.9267	0.9686
25	With unclassifiable category	Unclassifiable	Worried	Dominance	48.0743	71.8464	3.0143	5	0.6978	ns.	0.8322	0.9267	0.9686
25	With unclassifiable category	Unclassifiable	Worried	Uncertainty	23.44	31.4343	2.2091	5	0.8195	ns.	0.8322	0.9267	0.9686
25	With unclassifiable category	Unclassifiable	Worried	MBC:Cluster2	-79.2551	0	15.1802	5	0.0096	**	0.8322	0.9267	0.9686
25	With unclassifiable category	Unclassifiable	happy	Intercept	11.5988	21.8137	NA	NA	NA	NA	0.8322	0.9267	0.9686
25	With unclassifiable category	Unclassifiable	happy	Valence	51.1151	35.9746	10.3277	5	0.0665	ns.	0.8322	0.9267	0.9686
25	With unclassifiable category	Unclassifiable	happy	Arousal	-81.1032	6.9956	13.6047	5	0.0183	*	0.8322	0.9267	0.9686
25	With unclassifiable category	Unclassifiable	happy	Dominance	41.3657	51.8188	3.0143	5	0.6978	ns.	0.8322	0.9267	0.9686
25	With unclassifiable category	Unclassifiable	happy	Uncertainty	10.2232	24.1664	2.2091	5	0.8195	ns.	0.8322	0.9267	0.9686
25	With unclassifiable category	Unclassifiable	happy	MBC:Cluster2	14.3402	21.8137	15.1802	5	0.0096	**	0.8322	0.9267	0.9686
25	With unclassifiable category	Unclassifiable	relaxed	Intercept	128.5325	21.8343	NA	NA	NA	NA	0.8322	0.9267	0.9686
25	With unclassifiable category	Unclassifiable	relaxed	Valence	38.3931	35.9969	10.3277	5	0.0665	ns.	0.8322	0.9267	0.9686
25	With unclassifiable category	Unclassifiable	relaxed	Arousal	-89.4874	6.9784	13.6047	5	0.0183	*	0.8322	0.9267	0.9686
25	With unclassifiable category	Unclassifiable	relaxed	Dominance	43.7017	51.8222	3.0143	5	0.6978	ns.	0.8322	0.9267	0.9686
25	With unclassifiable category	Unclassifiable	relaxed	Uncertainty	-26.7236	24.1287	2.2091	5	0.8195	ns.	0.8322	0.9267	0.9686
25	With unclassifiable category	Unclassifiable	relaxed	MBC:Cluster2	-109.4489	21.8343	15.1802	5	0.0096	**	0.8322	0.9267	0.9686
25	With unclassifiable category	Unclassifiable	confused	Intercept	84.3123	76.5373	NA	NA	NA	NA	0.8322	0.9267	0.9686
25	With unclassifiable category	Unclassifiable	confused	Valence	96.0095	77.3942	10.3277	5	0.0665	ns.	0.8322	0.9267	0.9686
25	With unclassifiable category	Unclassifiable	confused	Arousal	24.9189	133.2414	13.6047	5	0.0183	*	0.8322	0.9267	0.9686
25	With unclassifiable category	Unclassifiable	confused	Dominance	-2.1107	97.9398	3.0143	5	0.6978	ns.	0.8322	0.9267	0.9686

ID	ModelType	Baseline	VECategory	TermInModel	Coefficient	StdErr	LRTest.ChiSq	DF	p	sig	McFadden	CoxSnell	Nagelkerke
25	With unclassifiable category	Unclassifiable	confused	Uncertainty	-30.6775	17.4255	2.2091	5	0.8195	ns.	0.8322	0.9267	0.9686
25	With unclassifiable category	Unclassifiable	confused	MBC:Cluster2	-42.3418	88.1753	15.1802	5	0.0096	**	0.8322	0.9267	0.9686
26	With unclassifiable category	Unclassifiable	admire	Intercept	27.8169	220.385	NA	NA	NA	NA	0.6746	0.8993	0.9303
26	With unclassifiable category	Unclassifiable	admire	Valence	-39.9488	40.6271	7.2492	5	0.2028	ns.	0.6746	0.8993	0.9303
26	With unclassifiable category	Unclassifiable	admire	Arousal	-29.146	36.7454	8.821	5	0.1164	ns.	0.6746	0.8993	0.9303
26	With unclassifiable category	Unclassifiable	admire	Dominance	26.3267	32.1504	12.9971	5	0.0234	*	0.6746	0.8993	0.9303
26	With unclassifiable category	Unclassifiable	admire	Uncertainty	-18.9802	358.9196	7.0683	5	0.2156	ns.	0.6746	0.8993	0.9303
26	With unclassifiable category	Unclassifiable	admire	MBC:Cluster2	-33.647	202.6273	8.8846	5	0.1138	ns.	0.6746	0.8993	0.9303
26	With unclassifiable category	Unclassifiable	discover	Intercept	4.413	512.5494	NA	NA	NA	NA	0.6746	0.8993	0.9303
26	With unclassifiable category	Unclassifiable	discover	Valence	-37.378	40.8534	7.2492	5	0.2028	ns.	0.6746	0.8993	0.9303
26	With unclassifiable category	Unclassifiable	discover	Arousal	-24.0385	37.4286	8.821	5	0.1164	ns.	0.6746	0.8993	0.9303
26	With unclassifiable category	Unclassifiable	discover	Dominance	33.3687	32.5405	12.9971	5	0.0234	*	0.6746	0.8993	0.9303
26	With unclassifiable category	Unclassifiable	discover	Uncertainty	31.92	351.6065	7.0683	5	0.2156	ns.	0.6746	0.8993	0.9303
26	With unclassifiable category	Unclassifiable	discover	MBC:Cluster2	-3.7238	509.097	8.8846	5	0.1138	ns.	0.6746	0.8993	0.9303
26	With unclassifiable category	Unclassifiable	enjoyable	Intercept	0.0346	71.417	NA	NA	NA	NA	0.6746	0.8993	0.9303
26	With unclassifiable category	Unclassifiable	enjoyable	Valence	-15.3443	38.1766	7.2492	5	0.2028	ns.	0.6746	0.8993	0.9303
26	With unclassifiable category	Unclassifiable	enjoyable	Arousal	-22.6812	37.4325	8.821	5	0.1164	ns.	0.6746	0.8993	0.9303
26	With unclassifiable category	Unclassifiable	enjoyable	Dominance	18.658	32.2984	12.9971	5	0.0234	*	0.6746	0.8993	0.9303
26	With unclassifiable category	Unclassifiable	enjoyable	Uncertainty	-0.425	661.2857	7.0683	5	0.2156	ns.	0.6746	0.8993	0.9303
26	With unclassifiable category	Unclassifiable	enjoyable	MBC:Cluster2	-11.216	71.417	8.8846	5	0.1138	ns.	0.6746	0.8993	0.9303
26	With unclassifiable category	Unclassifiable	explore	Intercept	-6.462	286.7234	NA	NA	NA	NA	0.6746	0.8993	0.9303
26	With unclassifiable category	Unclassifiable	explore	Valence	-26.3397	49.1483	7.2492	5	0.2028	ns.	0.6746	0.8993	0.9303
26	With unclassifiable category	Unclassifiable	explore	Arousal	7.1629	131.1869	8.821	5	0.1164	ns.	0.6746	0.8993	0.9303
26	With unclassifiable category	Unclassifiable	explore	Dominance	-22.1766	210.4987	12.9971	5	0.0234	*	0.6746	0.8993	0.9303
26	With unclassifiable category	Unclassifiable	explore	Uncertainty	33.4248	358.6578	7.0683	5	0.2156	ns.	0.6746	0.8993	0.9303
26	With unclassifiable category	Unclassifiable	explore	MBC:Cluster2	31.0841	340.5339	8.8846	5	0.1138	ns.	0.6746	0.8993	0.9303
26	With unclassifiable category	Unclassifiable	materialistic	Intercept	-21.3355	257.872	NA	NA	NA	NA	0.6746	0.8993	0.9303
26	With unclassifiable category	Unclassifiable	materialistic	Valence	-26.8487	49.2387	7.2492	5	0.2028	ns.	0.6746	0.8993	0.9303
26	With unclassifiable category	Unclassifiable	materialistic	Arousal	1.4261	131.4186	8.821	5	0.1164	ns.	0.6746	0.8993	0.9303
26	With unclassifiable category	Unclassifiable	materialistic	Dominance	-25.9307	210.5139	12.9971	5	0.0234	*	0.6746	0.8993	0.9303
26	With unclassifiable category	Unclassifiable	materialistic	Uncertainty	-42.0632	59.4152	7.0683	5	0.2156	ns.	0.6746	0.8993	0.9303
26	With unclassifiable category	Unclassifiable	materialistic	MBC:Cluster2	-5.6748	0	8.8846	5	0.1138	ns.	0.6746	0.8993	0.9303
27	With unclassifiable category	Unclassifiable	scary	Intercept	-24709.5537	0	NA	NA	NA	NA	1	0.9786	1
27	With unclassifiable category	Unclassifiable	scary	Valence	14367.641	0	25.905	6	0.0002	***	1	0.9786	1
27	With unclassifiable category	Unclassifiable	scary	Arousal	32860.2742	0	23.4353	6	0.0007	***	1	0.9786	1
27	With unclassifiable category	Unclassifiable	scary	Dominance	-7883.0674	0	32.6064	6	0	***	1	0.9786	1
27	With unclassifiable category	Unclassifiable	scary	Uncertainty	26892.0901	0	43.4567	6	0	***	1	0.9786	1
27	With unclassifiable category	Unclassifiable	scary	MBC:Cluster2	8664.6484	0	28.7143	6	0.0001	***	1	0.9786	1
27	With unclassifiable category	Unclassifiable	excited	Intercept	-16456.2165	0	NA	NA	NA	NA	1	0.9786	1
27	With unclassifiable category	Unclassifiable	excited	Valence	-9114.2992	0	25.905	6	0.0002	***	1	0.9786	1
27	With unclassifiable category	Unclassifiable	excited	Arousal	-7883.2905	0	23.4353	6	0.0007	***	1	0.9786	1
27	With unclassifiable category	Unclassifiable	excited	Dominance	-24268.625	0	32.6064	6	0	***	1	0.9786	1
27	With unclassifiable category	Unclassifiable	excited	Uncertainty	-113.7615	0	43.4567	6	0	***	1	0.9786	1
27	With unclassifiable category	Unclassifiable	excited	MBC:Cluster2	-8378.0932	0	28.7143	6	0.0001	***	1	0.9786	1
27	With unclassifiable category	Unclassifiable	Childlike	Intercept	-17225.7482	151.0165	NA	NA	NA	NA	1	0.9786	1
27	With unclassifiable category	Unclassifiable	Childlike	Valence	-16355.5456	894.9898	25.905	6	0.0002	***	1	0.9786	1
27	With unclassifiable category	Unclassifiable	Childlike	Arousal	-9094.6306	148.7059	23.4353	6	0.0007	***	1	0.9786	1
27	With unclassifiable category	Unclassifiable	Childlike	Dominance	14459.218	567.984	32.6064	6	0	***	1	0.9786	1
27	With unclassifiable category	Unclassifiable	Childlike	Uncertainty	-16225.6994	386.5322	43.4567	6	0	***	1	0.9786	1
27	With unclassifiable category	Unclassifiable	Childlike	MBC:Cluster2	13624.2291	151.0165	28.7143	6	0.0001	***	1	0.9786	1
27	With unclassifiable category	Unclassifiable	Free	Intercept	1168.5642	151.0165	NA	NA	NA	NA	1	0.9786	1
27	With unclassifiable category	Unclassifiable	Free	Valence	-15903.7068	894.9898	25.905	6	0.0002	***	1	0.9786	1

ID	ModelType	Baseline	VECategory	TermInModel	Coefficient	StdErr	LRTest.ChiSq	DF	p	sig	McFadden	CoxSnell	Nagelkerke
27	With unclassifiable category	Unclassifiable	Free	Arousal	-9120.7625	148.7059	23.4353	6	0.0007	***	1	0.9786	1
27	With unclassifiable category	Unclassifiable	Free	Dominance	14226.0048	567.984	32.6064	6	0	***	1	0.9786	1
27	With unclassifiable category	Unclassifiable	Free	Uncertainty	-12928.1546	386.5322	43.4567	6	0	***	1	0.9786	1
27	With unclassifiable category	Unclassifiable	Free	MBC:Cluster2	-4236.6269	151.0165	28.7143	6	0.0001	***	1	0.9786	1
27	With unclassifiable category	Unclassifiable	Helpless	Intercept	10163.0276	0	NA	NA	NA	NA	1	0.9786	1
27	With unclassifiable category	Unclassifiable	Helpless	Valence	-6191.4578	0	25.905	6	0.0002	***	1	0.9786	1
27	With unclassifiable category	Unclassifiable	Helpless	Arousal	-10204.5034	0	23.4353	6	0.0007	***	1	0.9786	1
27	With unclassifiable category	Unclassifiable	Helpless	Dominance	3941.277	0	32.6064	6	0	***	1	0.9786	1
27	With unclassifiable category	Unclassifiable	Helpless	Uncertainty	-24926.2919	0	43.4567	6	0	***	1	0.9786	1
27	With unclassifiable category	Unclassifiable	Helpless	MBC:Cluster2	-18894.3311	0	28.7143	6	0.0001	***	1	0.9786	1
27	With unclassifiable category	Unclassifiable	Lost	Intercept	23195.0582	137.38	NA	NA	NA	NA	1	0.9786	1
27	With unclassifiable category	Unclassifiable	Lost	Valence	-6687.7641	20.7861	25.905	6	0.0002	***	1	0.9786	1
27	With unclassifiable category	Unclassifiable	Lost	Arousal	-14292.6153	70.5646	23.4353	6	0.0007	***	1	0.9786	1
27	With unclassifiable category	Unclassifiable	Lost	Dominance	4788.5409	16.6822	32.6064	6	0	***	1	0.9786	1
27	With unclassifiable category	Unclassifiable	Lost	Uncertainty	15722.7467	25.7271	43.4567	6	0	***	1	0.9786	1
27	With unclassifiable category	Unclassifiable	Lost	MBC:Cluster2	-26010.1827	137.38	28.7143	6	0.0001	***	1	0.9786	1
28	With unclassifiable category	Unclassifiable	enjoyable	Intercept	6.4848	4.1055	NA	NA	NA	NA	0.5258	0.7836	0.8287
28	With unclassifiable category	Unclassifiable	enjoyable	Valence	11.7551	7.0824	10.2714	4	0.0361	*	0.5258	0.7836	0.8287
28	With unclassifiable category	Unclassifiable	enjoyable	Arousal	-1.0158	3.0519	3.9244	4	0.4163	ns.	0.5258	0.7836	0.8287
28	With unclassifiable category	Unclassifiable	enjoyable	Dominance	-6.3407	4.7678	7.815	4	0.0986	ns.	0.5258	0.7836	0.8287
28	With unclassifiable category	Unclassifiable	enjoyable	Uncertainty	-0.0635	1.2325	2.9037	4	0.5741	ns.	0.5258	0.7836	0.8287
28	With unclassifiable category	Unclassifiable	enjoyable	MBC:Cluster2	-14.4452	9.2438	6.6402	4	0.1562	ns.	0.5258	0.7836	0.8287
28	With unclassifiable category	Unclassifiable	bizzare	Intercept	-7.9852	104.4963	NA	NA	NA	NA	0.5258	0.7836	0.8287
28	With unclassifiable category	Unclassifiable	bizzare	Valence	-1.2071	3.2528	10.2714	4	0.0361	*	0.5258	0.7836	0.8287
28	With unclassifiable category	Unclassifiable	bizzare	Arousal	4.4844	3.3688	3.9244	4	0.4163	ns.	0.5258	0.7836	0.8287
28	With unclassifiable category	Unclassifiable	bizzare	Dominance	4.7511	3.6466	7.815	4	0.0986	ns.	0.5258	0.7836	0.8287
28	With unclassifiable category	Unclassifiable	bizzare	Uncertainty	-34.7351	451.5809	2.9037	4	0.5741	ns.	0.5258	0.7836	0.8287
28	With unclassifiable category	Unclassifiable	bizzare	MBC:Cluster2	-2.2722	6.3906	6.6402	4	0.1562	ns.	0.5258	0.7836	0.8287
28	With unclassifiable category	Unclassifiable	disgust	Intercept	-8.1698	30.6328	NA	NA	NA	NA	0.5258	0.7836	0.8287
28	With unclassifiable category	Unclassifiable	disgust	Valence	-0.3027	4.98	10.2714	4	0.0361	*	0.5258	0.7836	0.8287
28	With unclassifiable category	Unclassifiable	disgust	Arousal	-9.0917	17.1372	3.9244	4	0.4163	ns.	0.5258	0.7836	0.8287
28	With unclassifiable category	Unclassifiable	disgust	Dominance	-9.7393	20.3937	7.815	4	0.0986	ns.	0.5258	0.7836	0.8287
28	With unclassifiable category	Unclassifiable	disgust	Uncertainty	-37.9108	124.4615	2.9037	4	0.5741	ns.	0.5258	0.7836	0.8287
28	With unclassifiable category	Unclassifiable	disgust	MBC:Cluster2	-16.5987	60.5173	6.6402	4	0.1562	ns.	0.5258	0.7836	0.8287
28	With unclassifiable category	Unclassifiable	dislike	Intercept	-24.2209	37.8683	NA	NA	NA	NA	0.5258	0.7836	0.8287
28	With unclassifiable category	Unclassifiable	dislike	Valence	-0.7433	8.5508	10.2714	4	0.0361	*	0.5258	0.7836	0.8287
28	With unclassifiable category	Unclassifiable	dislike	Arousal	-0.7024	4.6744	3.9244	4	0.4163	ns.	0.5258	0.7836	0.8287
28	With unclassifiable category	Unclassifiable	dislike	Dominance	-1.9179	3.7231	7.815	4	0.0986	ns.	0.5258	0.7836	0.8287
28	With unclassifiable category	Unclassifiable	dislike	Uncertainty	-75.5956	118.5687	2.9037	4	0.5741	ns.	0.5258	0.7836	0.8287
28	With unclassifiable category	Unclassifiable	dislike	MBC:Cluster2	7.0958	20.8489	6.6402	4	0.1562	ns.	0.5258	0.7836	0.8287
29	Without unclassifiable category	phantasy curiosity	bored	Intercept	-4614.5523	169.6831	NA	NA	NA	NA	1	0.9891	1
29	Without unclassifiable category	phantasy curiosity	bored	Valence	-13229.556	283.3681	6.3146	10	0.7882	ns.	1	0.9891	1
29	Without unclassifiable category	phantasy curiosity	bored	Arousal	7617.6629	143.7722	0.8079	10	0.9999	ns.	1	0.9891	1
29	Without unclassifiable category	phantasy curiosity	bored	Dominance	13368.1902	265.518	19.1436	10	0.0385	*	1	0.9891	1
29	Without unclassifiable category	phantasy curiosity	bored	Uncertainty	1187.5672	40.731	15.5091	10	0.1146	ns.	1	0.9891	1
29	Without unclassifiable category	phantasy curiosity	bored	MBC:Cluster2	9953.6005	169.6831	0	10	1	ns.	1	0.9891	1

ID	ModelType	Baseline	VECategory	TermInModel	Coefficient	StdErr	LRTest.ChiSq	DF	p	sig	McFadden	CoxSnell	Nagelkerke
29	Without unclassifiable category	phantasy curiosity	stressed	Intercept	3753.4155	0	NA	NA	NA	NA	1	0.9891	1
29	Without unclassifiable category	phantasy curiosity	stressed	Valence	3941.858	0	6.3146	10	0.7882	ns.	1	0.9891	1
29	Without unclassifiable category	phantasy curiosity	stressed	Arousal	13517.968	0	0.8079	10	0.9999	ns.	1	0.9891	1
29	Without unclassifiable category	phantasy curiosity	stressed	Dominance	9107.6138	0	19.1436	10	0.0385	*	1	0.9891	1
29	Without unclassifiable category	phantasy curiosity	stressed	Uncertainty	-499.9171	0	15.5091	10	0.1146	ns.	1	0.9891	1
29	Without unclassifiable category	phantasy curiosity	stressed	MBC:Cluster2	-11949.5641	0	0	10	1	ns.	1	0.9891	1
29	Without unclassifiable category	phantasy curiosity	excited	Intercept	337.7236	4661.9117	NA	NA	NA	NA	1	0.9891	1
29	Without unclassifiable category	phantasy curiosity	excited	Valence	772.4646	22874.5779	6.3146	10	0.7882	ns.	1	0.9891	1
29	Without unclassifiable category	phantasy curiosity	excited	Arousal	170.762	3791.0232	0.8079	10	0.9999	ns.	1	0.9891	1
29	Without unclassifiable category	phantasy curiosity	excited	Dominance	-314.7203	9250.9005	19.1436	10	0.0385	*	1	0.9891	1
29	Without unclassifiable category	phantasy curiosity	excited	Uncertainty	491.1893	3986.4739	15.5091	10	0.1146	ns.	1	0.9891	1
29	Without unclassifiable category	phantasy curiosity	excited	MBC:Cluster2	-531.0715	4661.9117	0	10	1	ns.	1	0.9891	1
29	Without unclassifiable category	phantasy curiosity	peaceful	Intercept	-7855.6298	0	NA	NA	NA	NA	1	0.9891	1
29	Without unclassifiable category	phantasy curiosity	peaceful	Valence	6398.3438	0	6.3146	10	0.7882	ns.	1	0.9891	1
29	Without unclassifiable category	phantasy curiosity	peaceful	Arousal	-10318.5764	0	0.8079	10	0.9999	ns.	1	0.9891	1
29	Without unclassifiable category	phantasy curiosity	peaceful	Dominance	-14756.9914	0	19.1436	10	0.0385	*	1	0.9891	1
29	Without unclassifiable category	phantasy curiosity	peaceful	Uncertainty	-122.266	0	15.5091	10	0.1146	ns.	1	0.9891	1
29	Without unclassifiable category	phantasy curiosity	peaceful	MBC:Cluster2	-3507.3303	0	0	10	1	ns.	1	0.9891	1
29	Without unclassifiable category	phantasy curiosity	curious	Intercept	93.2048	503.4862	NA	NA	NA	NA	1	0.9891	1
29	Without unclassifiable category	phantasy curiosity	curious	Valence	586.9212	10.1972	6.3146	10	0.7882	ns.	1	0.9891	1
29	Without unclassifiable category	phantasy curiosity	curious	Arousal	1715.4003	363.2209	0.8079	10	0.9999	ns.	1	0.9891	1
29	Without unclassifiable category	phantasy curiosity	curious	Dominance	675.4221	2.6582	19.1436	10	0.0385	*	1	0.9891	1
29	Without unclassifiable category	phantasy curiosity	curious	Uncertainty	8937.246	40.3094	15.5091	10	0.1146	ns.	1	0.9891	1
29	Without unclassifiable category	phantasy curiosity	curious	MBC:Cluster2	1970.0396	503.4862	0	10	1	ns.	1	0.9891	1
29	Without unclassifiable category	phantasy curiosity	intrigued	Intercept	4264.137	0	NA	NA	NA	NA	1	0.9891	1
29	Without unclassifiable category	phantasy curiosity	intrigued	Valence	-3538.2667	0	6.3146	10	0.7882	ns.	1	0.9891	1

ID	ModelType	Baseline	VECategory	TermInModel	Coefficient	StdErr	LRTest.ChiSq	DF	p	sig	McFadden	CoxSnell	Nagelkerke
29	Without unclassifiable category	phantasy curiosity	intrigued	Arousal	-8717.8067	0	0.8079	10	0.9999	ns.	1	0.9891	1
29	Without unclassifiable category	phantasy curiosity	intrigued	Dominance	-8800.2758	0	19.1436	10	0.0385	*	1	0.9891	1
29	Without unclassifiable category	phantasy curiosity	intrigued	Uncertainty	-1690.1053	0	15.5091	10	0.1146	ns.	1	0.9891	1
29	Without unclassifiable category	phantasy curiosity	intrigued	MBC:Cluster2	-11506.6523	0	0	10	1	ns.	1	0.9891	1
29	Without unclassifiable category	phantasy curiosity	terrified	Intercept	-18294.0954	0	NA	NA	NA	NA	1	0.9891	1
29	Without unclassifiable category	phantasy curiosity	terrified	Valence	-2838.7949	0	6.3146	10	0.7882	ns.	1	0.9891	1
29	Without unclassifiable category	phantasy curiosity	terrified	Arousal	10915.6201	0	0.8079	10	0.9999	ns.	1	0.9891	1
29	Without unclassifiable category	phantasy curiosity	terrified	Dominance	-6027.1259	0	19.1436	10	0.0385	*	1	0.9891	1
29	Without unclassifiable category	phantasy curiosity	terrified	Uncertainty	4221.0341	0	15.5091	10	0.1146	ns.	1	0.9891	1
29	Without unclassifiable category	phantasy curiosity	terrified	MBC:Cluster2	156.9008	0	0	10	1	ns.	1	0.9891	1
29	Without unclassifiable category	phantasy curiosity	degrading	Intercept	-1627.1978	0	NA	NA	NA	NA	1	0.9891	1
29	Without unclassifiable category	phantasy curiosity	degrading	Valence	-4624.7246	0	6.3146	10	0.7882	ns.	1	0.9891	1
29	Without unclassifiable category	phantasy curiosity	degrading	Arousal	-27.9762	0	0.8079	10	0.9999	ns.	1	0.9891	1
29	Without unclassifiable category	phantasy curiosity	degrading	Dominance	-5265.5332	0	19.1436	10	0.0385	*	1	0.9891	1
29	Without unclassifiable category	phantasy curiosity	degrading	Uncertainty	-2390.7583	0	15.5091	10	0.1146	ns.	1	0.9891	1
29	Without unclassifiable category	phantasy curiosity	degrading	MBC:Cluster2	-7078.9294	0	0	10	1	ns.	1	0.9891	1
29	Without unclassifiable category	phantasy curiosity	lonely	Intercept	2045.6179	0	NA	NA	NA	NA	1	0.9891	1
29	Without unclassifiable category	phantasy curiosity	lonely	Valence	-11063.8378	0	6.3146	10	0.7882	ns.	1	0.9891	1
29	Without unclassifiable category	phantasy curiosity	lonely	Arousal	10000.9946	0	0.8079	10	0.9999	ns.	1	0.9891	1
29	Without unclassifiable category	phantasy curiosity	lonely	Dominance	19714.0012	0	19.1436	10	0.0385	*	1	0.9891	1
29	Without unclassifiable category	phantasy curiosity	lonely	Uncertainty	-900.4318	0	15.5091	10	0.1146	ns.	1	0.9891	1
29	Without unclassifiable category	phantasy curiosity	lonely	MBC:Cluster2	-6679.6024	0	0	10	1	ns.	1	0.9891	1
29	Without unclassifiable category	phantasy curiosity	weird	Intercept	-4066.9182	0	NA	NA	NA	NA	1	0.9891	1
29	Without unclassifiable category	phantasy curiosity	weird	Valence	10233.6177	0	6.3146	10	0.7882	ns.	1	0.9891	1
29	Without unclassifiable category	phantasy curiosity	weird	Arousal	16071.8876	0	0.8079	10	0.9999	ns.	1	0.9891	1
29	Without unclassifiable category	phantasy curiosity	weird	Dominance	-2307.8259	0	19.1436	10	0.0385	*	1	0.9891	1

ID	ModelType	Baseline	VECategory	TermInModel	Coefficient	StdErr	LRTest.ChiSq	DF	p	sig	McFadden	CoxSnell	Nagelkerke
29	Without unclassifiable category	phantasy curiosity	weird	Uncertainty	1204.0029	0	15.5091	10	0.1146	ns.	1	0.9891	1
29	Without unclassifiable category	phantasy curiosity	weird	MBC:Cluster2	-3343.754	0	0	10	1	ns.	1	0.9891	1
30	Without unclassifiable category	relaxed	scary	Intercept	-198.9848	1.2638	NA	NA	NA	NA	0.7831	0.9198	0.958
30	Without unclassifiable category	relaxed	scary	Valence	77.9505	1.0013	-0.1559	5	1	ns.	0.7831	0.9198	0.958
30	Without unclassifiable category	relaxed	scary	Arousal	272.4214	1.9329	1.3138	5	0.9335	ns.	0.7831	0.9198	0.958
30	Without unclassifiable category	relaxed	scary	Dominance	-112.8065	1.488	6.9364	5	0.2254	ns.	0.7831	0.9198	0.958
30	Without unclassifiable category	relaxed	scary	Uncertainty	40.4788	0.2898	2.3331	5	0.8014	ns.	0.7831	0.9198	0.958
30	Without unclassifiable category	relaxed	scary	MBC:Cluster2	56.4116	0	1.9641	5	0.8541	ns.	0.7831	0.9198	0.958
30	Without unclassifiable category	relaxed	happy	Intercept	-61.1354	2.9571	NA	NA	NA	NA	0.7831	0.9198	0.958
30	Without unclassifiable category	relaxed	happy	Valence	-1.4939	3.2012	-0.1559	5	1	ns.	0.7831	0.9198	0.958
30	Without unclassifiable category	relaxed	happy	Arousal	1.8317	3.1571	1.3138	5	0.9335	ns.	0.7831	0.9198	0.958
30	Without unclassifiable category	relaxed	happy	Dominance	-1.1197	2.0269	6.9364	5	0.2254	ns.	0.7831	0.9198	0.958
30	Without unclassifiable category	relaxed	happy	Uncertainty	-25.6598	29.6923	2.3331	5	0.8014	ns.	0.7831	0.9198	0.958
30	Without unclassifiable category	relaxed	happy	MBC:Cluster2	58.1316	2.9571	1.9641	5	0.8541	ns.	0.7831	0.9198	0.958
30	Without unclassifiable category	relaxed	Uneasy	Intercept	172.7428	63.1474	NA	NA	NA	NA	0.7831	0.9198	0.958
30	Without unclassifiable category	relaxed	Uneasy	Valence	-4.5674	40.8047	-0.1559	5	1	ns.	0.7831	0.9198	0.958
30	Without unclassifiable category	relaxed	Uneasy	Arousal	4.5969	17.3359	1.3138	5	0.9335	ns.	0.7831	0.9198	0.958
30	Without unclassifiable category	relaxed	Uneasy	Dominance	-8.5205	20.8102	6.9364	5	0.2254	ns.	0.7831	0.9198	0.958
30	Without unclassifiable category	relaxed	Uneasy	Uncertainty	-137.6145	366.6763	2.3331	5	0.8014	ns.	0.7831	0.9198	0.958
30	Without unclassifiable category	relaxed	Uneasy	MBC:Cluster2	-195.5995	82.8071	1.9641	5	0.8541	ns.	0.7831	0.9198	0.958
30	Without unclassifiable category	relaxed	uncomfortable	Intercept	-42.1001	132.801	NA	NA	NA	NA	0.7831	0.9198	0.958
30	Without unclassifiable category	relaxed	uncomfortable	Valence	-37.8478	91.3281	-0.1559	5	1	ns.	0.7831	0.9198	0.958
30	Without unclassifiable category	relaxed	uncomfortable	Arousal	4.6662	315.4763	1.3138	5	0.9335	ns.	0.7831	0.9198	0.958
30	Without unclassifiable category	relaxed	uncomfortable	Dominance	-218.525	36.0992	6.9364	5	0.2254	ns.	0.7831	0.9198	0.958
30	Without unclassifiable category	relaxed	uncomfortable	Uncertainty	-40.2727	187.9303	2.3331	5	0.8014	ns.	0.7831	0.9198	0.958
30	Without unclassifiable category	relaxed	uncomfortable	MBC:Cluster2	-70.847	0	1.9641	5	0.8541	ns.	0.7831	0.9198	0.958
30	Without unclassifiable category	relaxed	Sad	Intercept	-20.5546	NA	NA	NA	NA	NA	0.7831	0.9198	0.958
30	Without unclassifiable category	relaxed	Sad	Valence	-52.8374	0	-0.1559	5	1	ns.	0.7831	0.9198	0.958
30	Without unclassifiable category	relaxed	Sad	Arousal	54.9535	0	1.3138	5	0.9335	ns.	0.7831	0.9198	0.958
30	Without unclassifiable category	relaxed	Sad	Dominance	-51.3924	NA	6.9364	5	0.2254	ns.	0.7831	0.9198	0.958
30	Without unclassifiable category	relaxed	Sad	Uncertainty	5.8841	0	2.3331	5	0.8014	ns.	0.7831	0.9198	0.958
30	Without unclassifiable category	relaxed	Sad	MBC:Cluster2	-35.9152	0	1.9641	5	0.8541	ns.	0.7831	0.9198	0.958

F.4 R session information

```
R version 3.4.0 (2017-04-21)
Platform: x86_64-pc-linux-gnu (64-bit)
Running under: Ubuntu 16.04.2 LTS

Matrix products: default
BLAS: /usr/lib/libblas/libblas.so.3.6.0
LAPACK: /usr/lib/lapack/liblapack.so.3.6.0

locale:
 [1] LC_CTYPE=en_GB.UTF-8       LC_NUMERIC=C               LC_TIME=en_GB.UTF-8        LC_COLLATE=en_GB.UTF-8
 [5] LC_MONETARY=en_GB.UTF-8    LC_MESSAGES=en_GB.UTF-8    LC_PAPER=en_GB.UTF-8       LC_NAME=C
 [9] LC_ADDRESS=C               LC_TELEPHONE=C             LC_MEASUREMENT=en_GB.UTF-8 LC_IDENTIFICATION=C

attached base packages:
[1] grid      stats     graphics  grDevices utils     datasets  methods   base

other attached packages:
 [1] nnet_7.3-12         vcd_1.4-3           SnowballC_0.5.1     rlist_0.4.6.1       tidyr_0.6.3
 [6] psych_1.7.5         texreg_1.36.23      xtable_1.8-2        ggdendro_0.1-20     ggrepel_0.6.5
[11] ggplot2_2.2.1       lme4_1.1-13         Matrix_1.2-10       mclust_5.3          poLCA_1.4.1
[16] MASS_7.3-47         scatterplot3d_0.3-40 clusteval_0.1      cooccur_1.3         rapportools_1.0
[21] reshape_0.8.6       plyr_1.8.4          stringr_1.2.0       data.table_1.10.4

loaded via a namespace (and not attached):
 [1] nlme_3.1-131        matrixStats_0.52.2  pbkrtest_0.4-7      xts_0.10-0          threejs_0.3.1
 [6] rstan_2.16.2        tools_3.4.0         R6_2.2.1            DT_0.2              lazyeval_0.2.0
[11] mgcv_1.8-17         colorspace_1.3-2    gridExtra_2.2.1     mnormt_1.5-5        Brobdingnag_1.2-4
```

```
[16] compiler_3.4.0     quantreg_5.33        BiasedUrn_1.07       SparseM_1.77         shinyjs_0.9.1
[21] colourpicker_0.3   scales_0.4.1         dygraphs_1.1.1.4     lmtest_0.9-35        brms_1.9.0
[26] mvtnorm_1.0-6      digest_0.6.12        StanHeaders_2.16.0-1 foreign_0.8-67       minqa_1.2.4
[31] base64enc_0.1-3    pkgconfig_2.0.1      htmltools_0.3.6      htmlwidgets_0.9      rlang_0.1
[36] shiny_1.0.3        bindr_0.1            zoo_1.8-0            crosstalk_1.0.0      gtools_3.5.0
[41] dplyr_0.7.2        car_2.1-5            inline_0.3.14        magrittr_1.5         loo_1.1.0
[46] bayesplot_1.3.0    Rcpp_0.12.10         munsell_0.4.3        abind_1.4-5          stringi_1.1.5
[51] parallel_3.4.0     miniUI_0.1.1         lattice_0.20-35      splines_3.4.0        pander_0.6.1
[56] epiR_0.9-87        generalhoslem_1.2.5  igraph_1.1.2         markdown_0.8         shinystan_2.4.0
[61] reshape2_1.4.2     stats4_3.4.0         rstantools_1.3.0     glue_1.1.1           nloptr_1.0.4
[66] httpuv_1.3.5       MatrixModels_0.4-1   gtable_0.2.0         assertthat_0.2.0     mime_0.5
[71] coda_0.19-1        survival_2.41-3      rsconnect_0.8.5      tibble_1.3.1         shinythemes_1.1.1
[76] bindrcpp_0.2       gmp_0.5-13.1         bridgesampling_0.2-2
```

Appendix G

Real-world PAD ratings

G.1 Instruction pack for the Android app

RATING
EMOTIONAL
EXPERIENCES

Information pack for participants. Please read carefully!

Researcher: Caterina Constantinescu
Supervisors: Dr. Sarah E. MacPherson & Dr. Adam Moore

Abstract

Thank you for agreeing to participate in this study! You will be beeped 4 times a day over 2 weeks, and asked to rate any emotional experiences you've just had. For instance, if you happen to be enjoying a delicious meal, or feel under the weather because of a cold, you can describe & rate these experiences when using our app.

You will be beeped usually between 10am and 10pm (including weekends), with a minimum gap of 2h between beeps. If a beep occurs at an inconvenient time, you can postpone the rating procedure for 30 minutes, but **please do not abuse this feature**.

Your participant number: ....

Your participant ID in the app: ....

Your study start date (next day after installing app): ....

Your study end date: ....

1 Researcher contact details

Should anything go wrong with the app or if you simply need more information:

- **Caterina Constantinescu (Researcher)**: caterina.constantinescu@ed.ac.uk

- Dr. Sarah E. MacPherson (Supervisor): sarah.macpherson@ed.ac.uk
- Dr. Adam Moore (Supervisor): amoore23@exseed.ed.ac.uk

2 Pre-requisites

To be able to participate in this study, you **MUST**:

- Have an Android phone;
- Have access to wifi / signal & enough data when filling in the survey questions, in order to send your responses to the researcher.

3 Study overview, duration, and payment

The study will involve:

- Answering a few questionnaires asking about your mood & lifestyle (5-10 mins), in the researcher's lab.
- Seeing how to install & use the app, and setting up your Participant ID.
- Starting from the following day, your phone will notify you when you should provide your ratings, at random intervals throughout the day (usually between 10am - 10pm).

Details about payment:

- Each individual rating session takes very little time, so you will be reimbursed £15 for your overall time commitment (including the questionnaires from the beginning), at the end of the study.
- You will get a chance to choose whether to receive this payment through an Amazon voucher by email, or cash if you return to the lab at the end of the study.

Your rights:

- You are free to stop the experiment at any point if you so wish.
- The data we collect through the app do not contain any personally identifiable information, so your anonymity is guaranteed.

4 Instructions for installing the app

1. Go to **Settings** ▷ **Security** ▷ **Unknown sources**, on your Android phone.
2. Enable the installation of apps from unkown sources. Our app is safe, with no malicious content, so don't worry.
3. Please visit the link: http://bit.ly/RatingEmotions and tap the Download symbol (top right corner). Pick **"Direct download"**, and then accept the `android-debug.apk` file.
4. Once the download is complete, please tap the app to install it.
5. Return to your Settings menu and remember to **switch off the "Install from unknown sources" option**.
6. Finally, navigate to the app in your phone menu, and start it up. You can look around for the app's logo, shown in Figure 1.
7. You will now be able to generate your participant ID by following the instructions on-screen.
8. Once you are shown that the data has been successfully sent to our servers, you can close the app from the **Recent Applications** button, shown in Figure 2 - otherwise the app is left in stand-by and may not function normally.

Figure 1: App logo

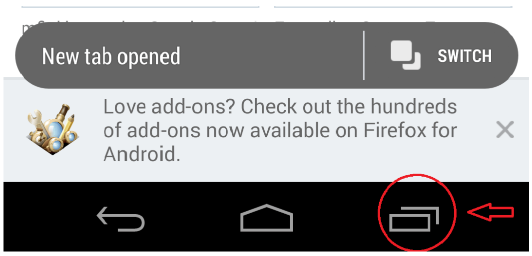
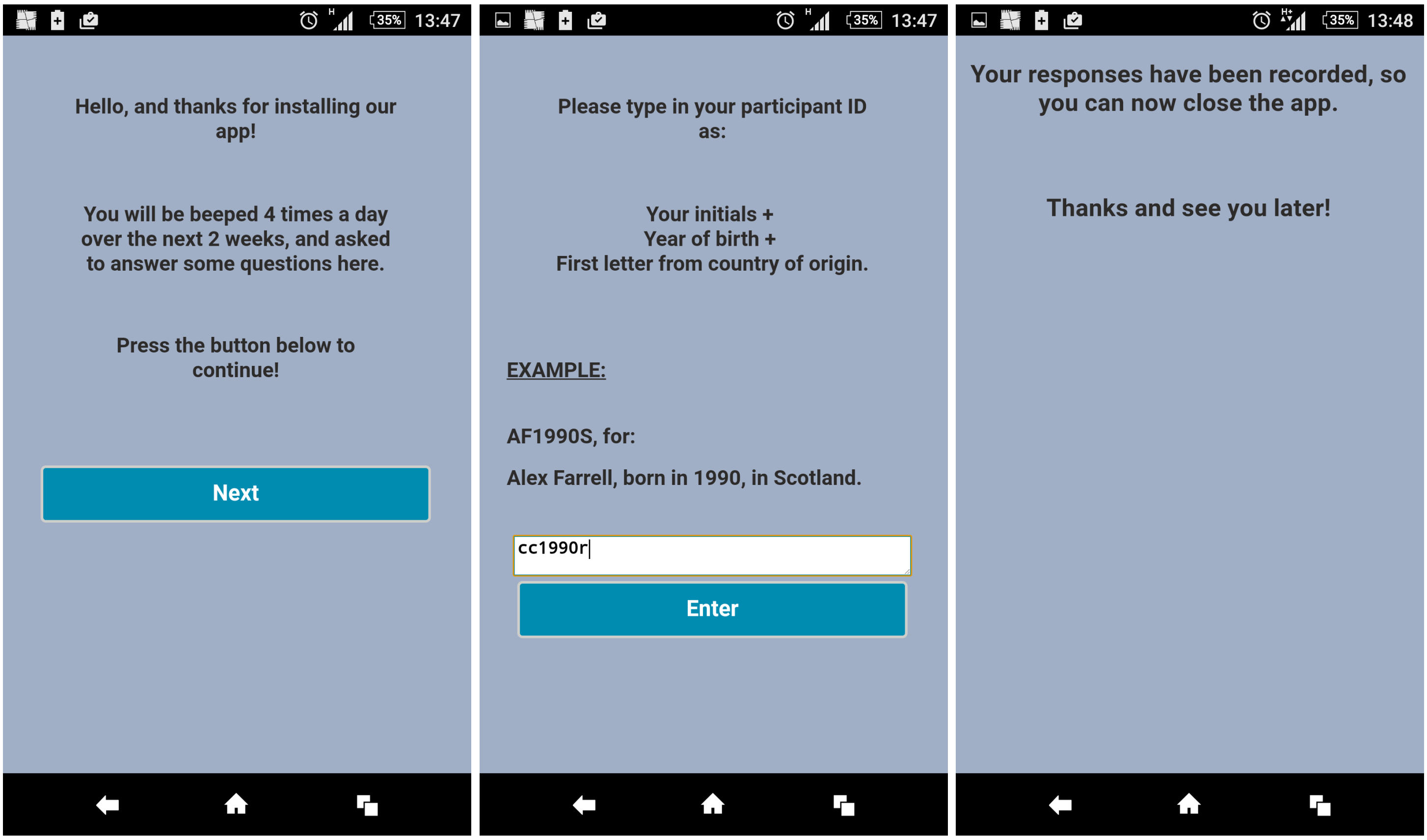


Figure 2: Recent Applications button, marked with an arrow

5 Setting your Participant ID

The first screens you will see on the app are shown in Figure 3. Your main task here is to generate your Participant ID, according to the instructions on your screen. Once this is done, your notification schedule will have been created, and you are set to start providing ratings from the following morning (some time around 10am).

Figure 3: First startup of the app



(a) Welcome screen

(b) Creating your participant ID

(c) Information sent to server. Please only ever close the app **after** you see this message.

6 Rating scales

Each time you receive a notification from the app, you'll be asked to rate any emotional experiences you were having / had just had within the last 30 mins. You'll be asked to provide ratings using the three scales from Figure 4:

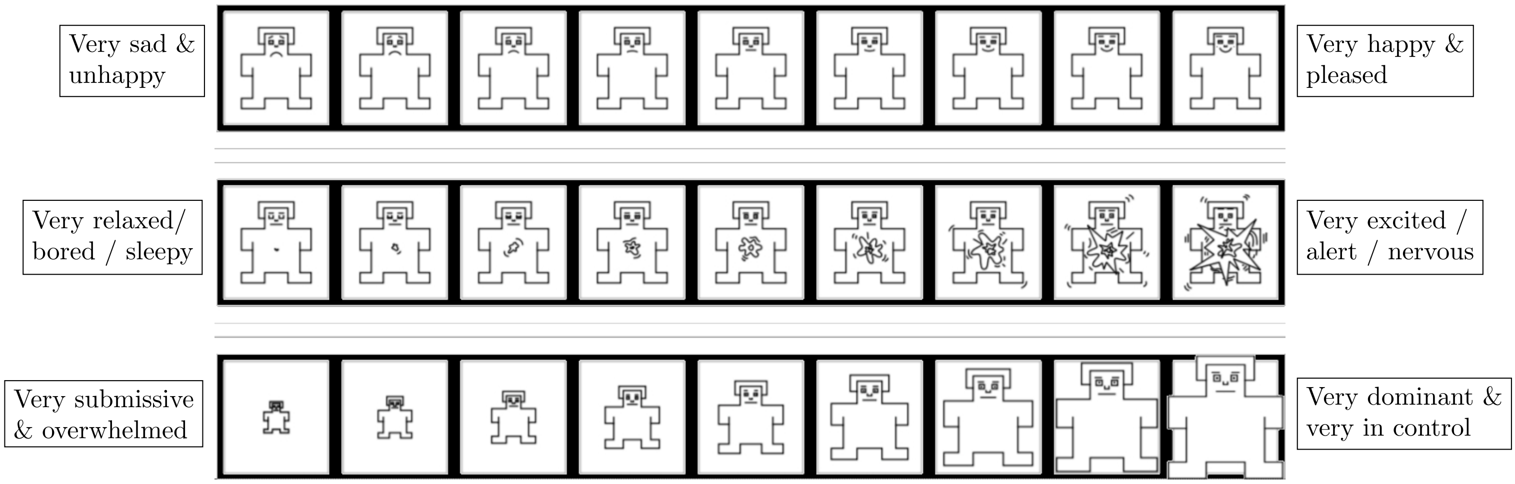
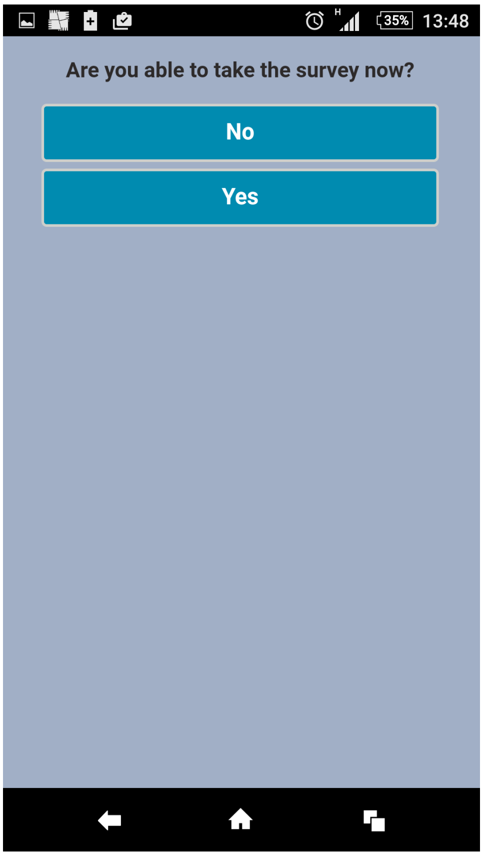


Figure 4: Three of the rating scales used in the app. **Example** for third row: If petting a kitten, you'd probably feel dominant / in control of the situation, whereas if being chased by a grizzly bear, you'd feel submissive and overwhelmed.

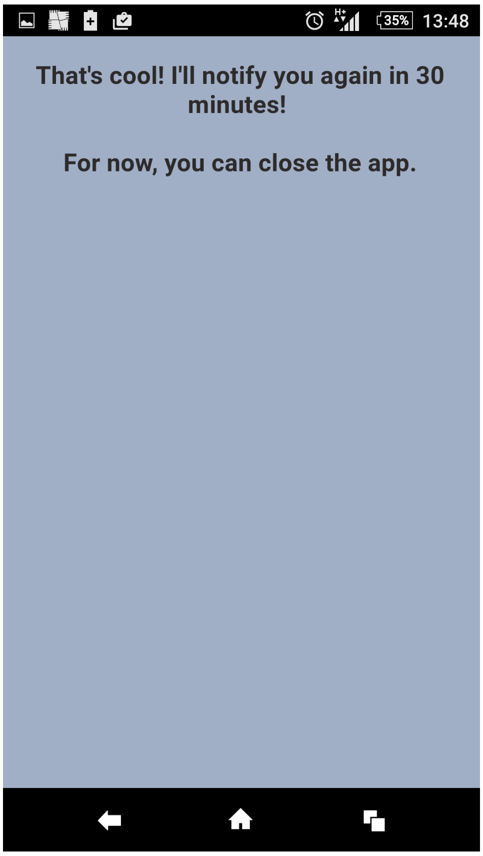
7 Providing daily ratings

After creating your Participant ID, you will receive your first app notification on the following day, some time around 10am. **Notifications will appear as a small bell at the top of your screen**. Take a look at Figures 5 to 10 to see what you will be asked.

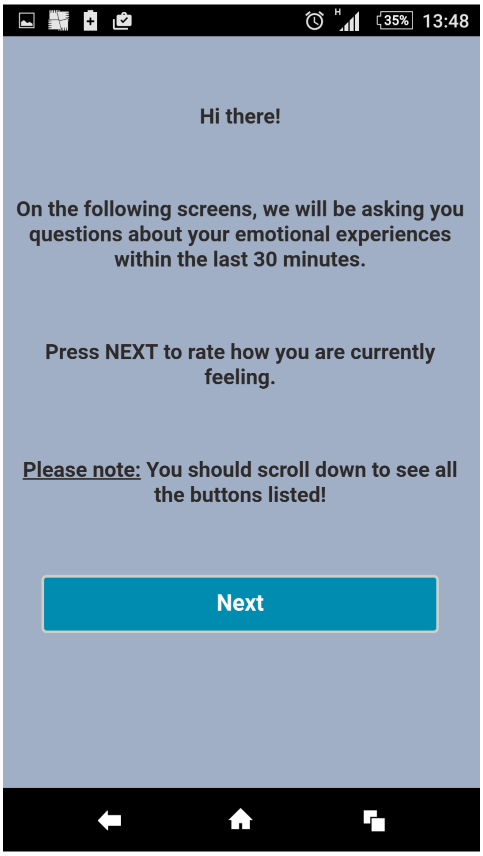
Figure 5: Daily notifications / ratings from the app. **Please do not abuse the snooze function.**



(a) To snooze or not to snooze? That is the question.

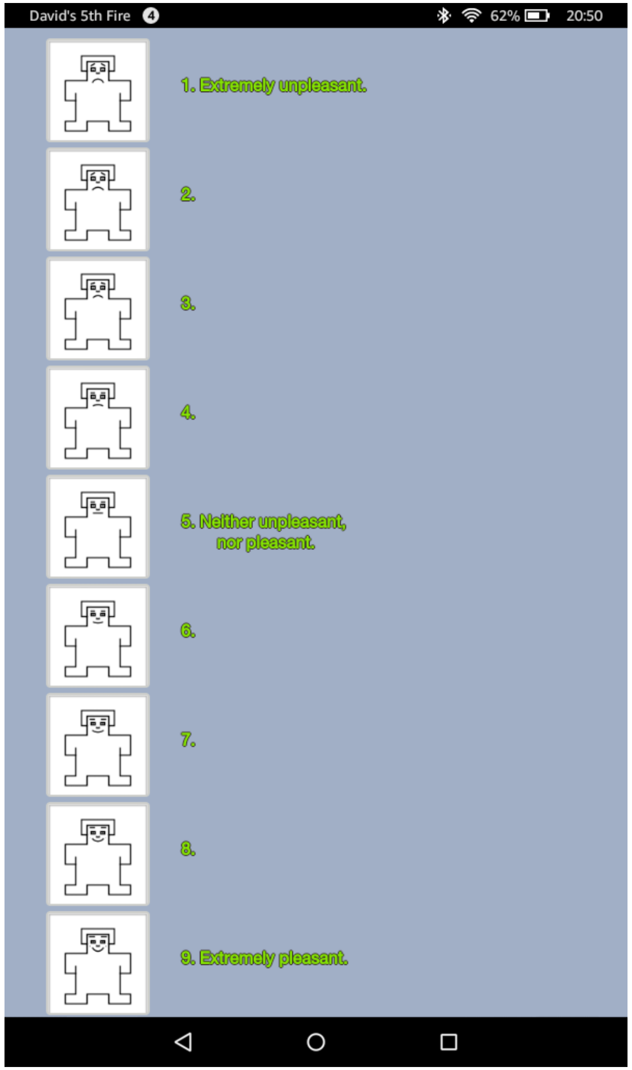


(b) Snoozed, if now is not a good time, OR:

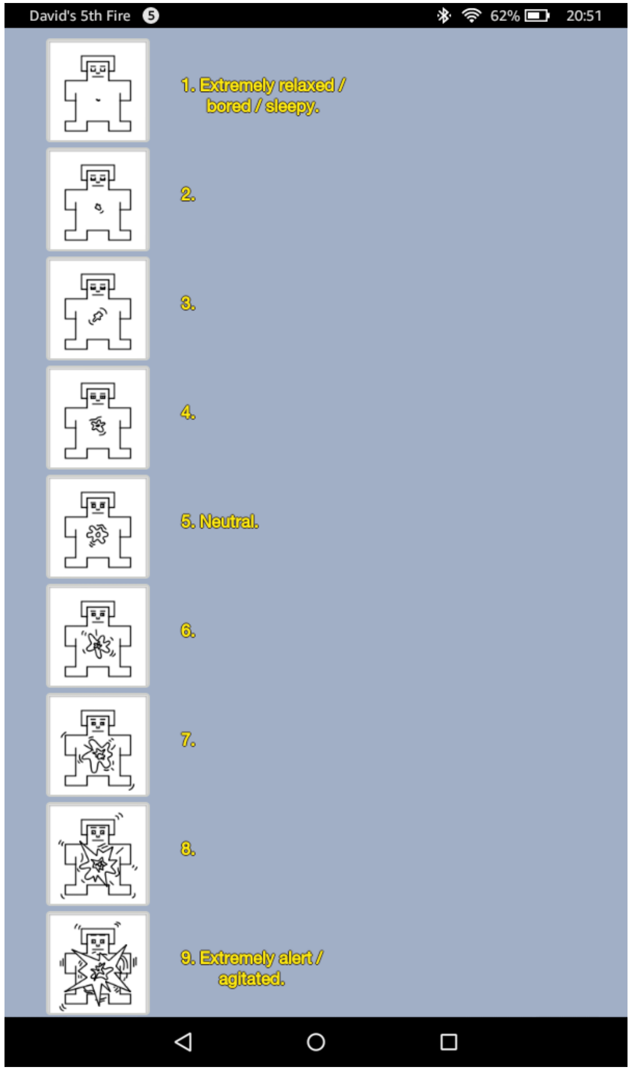


(c) Not snoozed, now is fine!

Figure 6: App ratings (I). Pick one of the 9 images from each screen / scale to indicate how you are feeling. **Depending on the size of your phone screen, you might need to scroll down to see all 9 images from each scale!**



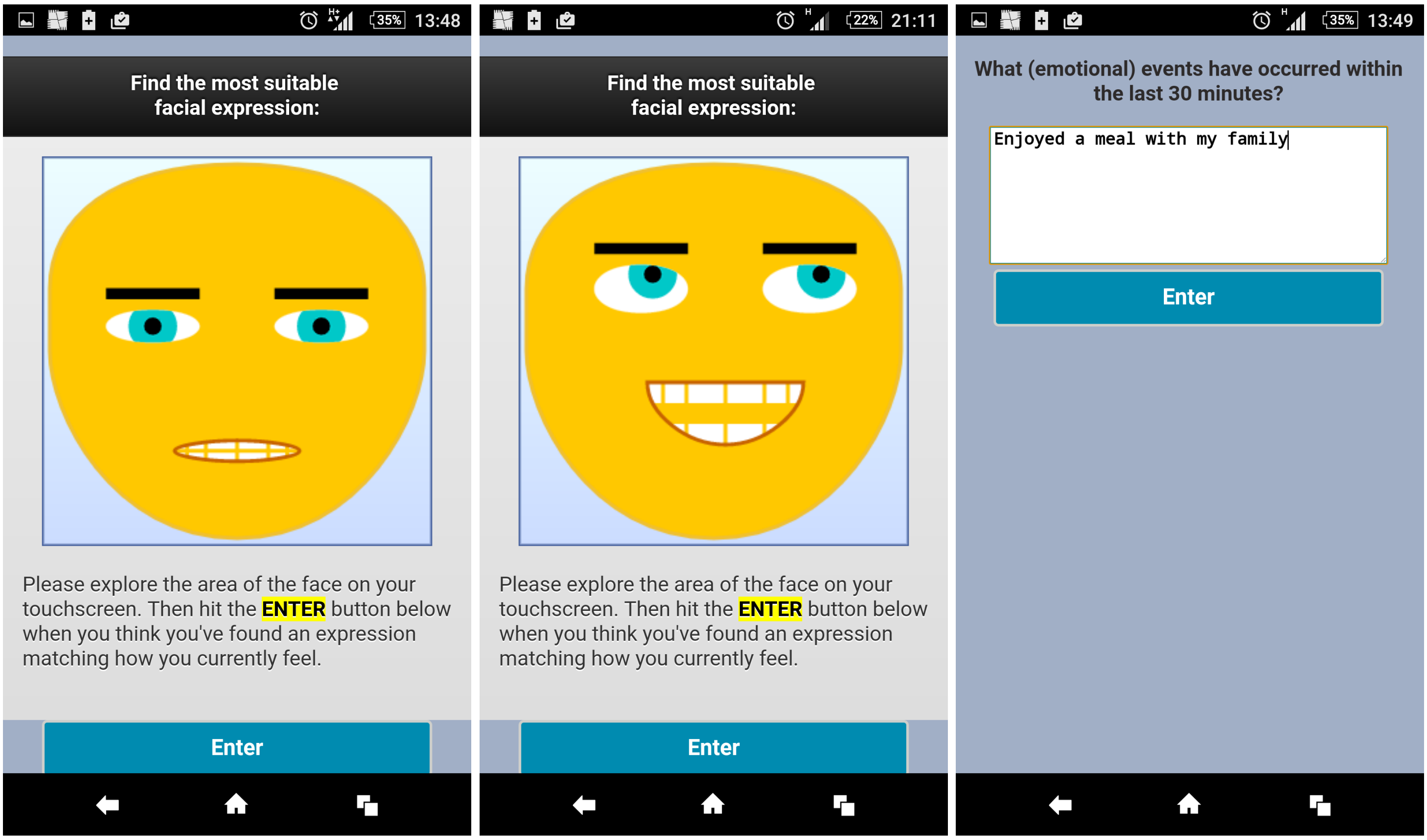
(a) Sad - Happy



(b) Relaxed / Bored / Sleepy - Excited / Alert / Nervous

(c) Overwhelmed / Submissive - In control / Dominant

Figure 7: App ratings (II). **Please wait to load the full screen below (you will need an Internet connection for this, and to wait a few seconds).** Then, find a facial expression that matches how you feel, by dragging your finger over the face. Once you think you have found a suitable expression, click the Enter button below. Next, you will be asked to write a short message about what emotional event has just happened / is happening currently.

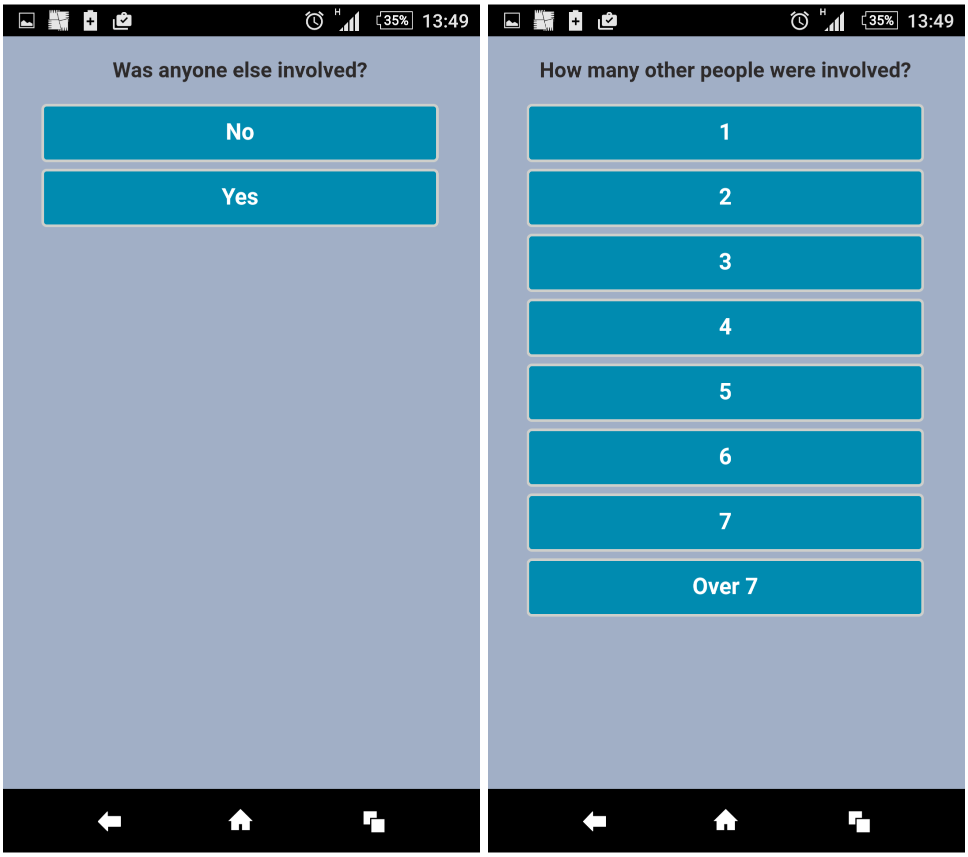


(a) Affect face - default expression

(b) Affect face - dragging to create a facial expression

(c) Describing the emotional situation: **ideally, 1-3 lines of text**

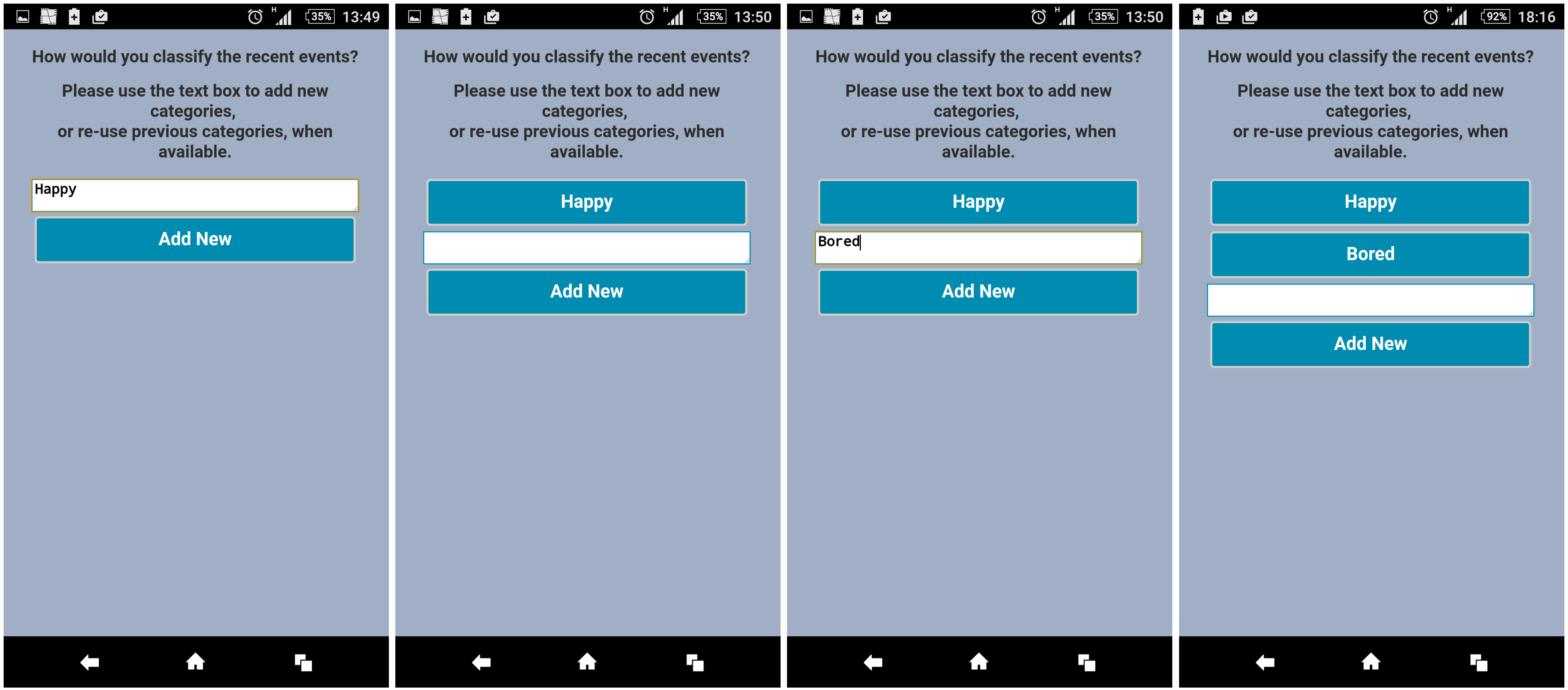
Figure 8: App ratings (III). These questions aim to find out whether the emotional event involved other people as well. If yes, you are asked how many. If not, you will skip to the next question.



(a) Any other people involved? If so, you will see the next screen as well.

(b) How many other people?

Figure 9: App ratings (IV). Here you will be asked to create your own categories for the emotional events you have just experienced. Please do **not** create one category label for every event. Rather, try to figure out if there is a pattern across all the events, and try to group them accordingly.



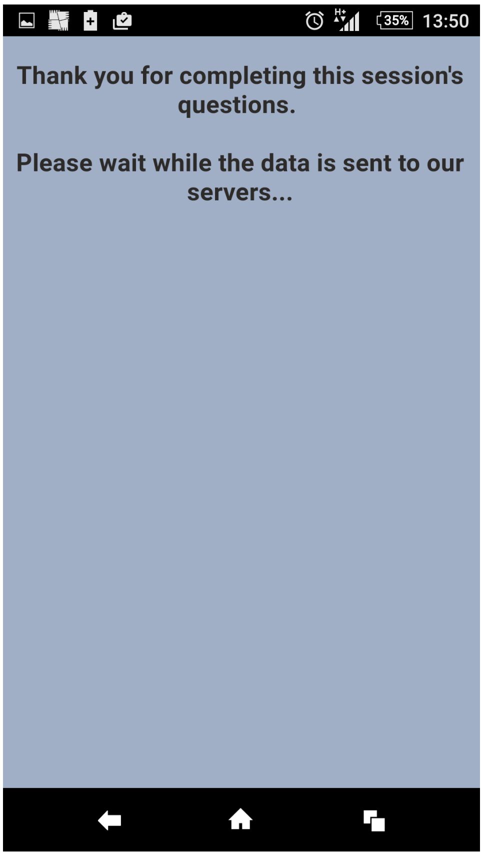
(a) Naming your first category "happy" if it seems appropriate for the event.

(b) The next time you resume the app, "happy" will already be listed as a category ready to use.

(c) In case "happy" is not suitable this time, we can create another category this time round.

(d) The next time you use your app, the "bored" category will also be displayed, and ready for use.

Figure 10: Final messages for each session.



(a) Data sent to server. This usually takes a few seconds if you have signal.

(b) Data was sent and you can now close the app. **Please only close the app AFTER you see this message. Then, to close it, please use the Recent Apps button (shown in Figure 2)**

8 Consent form

Participant number: .... Participant ID in app: ....

By signing below, you are agreeing that:

1. you have read and understood the Information Pack,
2. questions about your participation in this study have been answered satisfactorily,
3. you are taking part in this research study voluntarily (without coercion) and understand you can quit at any time,
4. you understand that participation in this study involves completion of some standardised tests [Toronto Alexithymia Scale-20, Patient Health Questionnaire-2] which are used as preliminary screens for clinical conditions of which you may not be aware. Scores from these tests would not be sufficient basis for clinical decisions or diagnosis, contain substantial margins of error, and are not used for diagnostic purposes in this study. Though it is not possible to provide feedback of individual scores to participants, these scores might hint at health problems. In the event that you produce scores of potential clinical concern, researchers should (pick **one** box):

☐ Contact me at : ....

☐ Contact my GP at: ....

☐ Do nothing. I absolve the researchers of any obligation to contact me about this.

....	....	....	....
Participant's name (can also use just the initials)	Email	Signature	Date

Caterina Constantinescu
(Person obtaining consent)

.... Signature

Payment preference (circle one): Amazon voucher via email provided above / Collect cash after completing the study

G.2 R session information

```
R version 3.4.1 (2017-06-30)
Platform: x86_64-pc-linux-gnu (64-bit)
Running under: Ubuntu 16.04.3 LTS

Matrix products: default
BLAS: /usr/lib/libblas/libblas.so.3.6.0
LAPACK: /usr/lib/lapack/liblapack.so.3.6.0

locale:
 [1] LC_CTYPE=en_GB.UTF-8       LC_NUMERIC=C               LC_TIME=en_GB.UTF-8        LC_COLLATE=en_GB.UTF-8
 [5] LC_MONETARY=en_GB.UTF-8    LC_MESSAGES=en_GB.UTF-8    LC_PAPER=en_GB.UTF-8       LC_NAME=en_GB.UTF-8
 [9] LC_ADDRESS=en_GB.UTF-8     LC_TELEPHONE=en_GB.UTF-8   LC_MEASUREMENT=en_GB.UTF-8 LC_IDENTIFICATION=en_GB.UTF-8

attached base packages:
[1] stats     graphics  grDevices utils     datasets  methods   base

other attached packages:
 [1] texreg_1.36.23             xtable_1.8-2               kSamples_1.2-7             SuppDists_1.1-9.4
 [5] Matching_4.9-2             MASS_7.3-47                lmerTest_2.0-33            LMERConvenienceFunctions_2.10
 [9] lme4_1.1-13                Matrix_1.2-11              mclust_5.3                 psych_1.7.5
[13] gridExtra_2.2.1            ggrepel_0.6.5              anytime_0.2.2              chron_2.3-50
[17] tables_0.8                 Hmisc_4.0-3                ggplot2_2.2.1              Formula_1.2-2
[21] survival_2.41-3            lattice_0.20-35            plyr_1.8.4                 splitstackshape_1.4.2
[25] broom_0.4.2                tidyr_0.6.3                dplyr_0.7.2                stringr_1.2.0
[29] XLConnect_0.2-13           XLConnectJars_0.2-13       googlesheets_0.2.2         statip_0.1.4
[33] topicmodels_0.2-6          koRpus_0.10-2              data.table_1.10.4
```

```
loaded via a namespace (and not attached):
 [1] maps_3.2.0          jsonlite_1.4        splines_3.4.1       dotCall64_0.9-04    shiny_1.0.3         assertthat_0.2.0
 [7] stats4_3.4.1        latticeExtra_0.6-28 cellranger_1.1.0    slam_0.1-40         backports_1.1.0     glue_1.1.1
[13] digest_0.6.12       RColorBrewer_1.1-2  checkmate_1.8.3     minqa_1.2.4         colorspace_1.3-2    httpuv_1.3.5
[19] htmltools_0.3.6     tm_0.7-1            pkgconfig_2.0.1     purrr_0.2.2.2       scales_0.4.1        htmlTable_1.9
[25] tibble_1.3.1        mgcv_1.8-19         nnet_7.3-12         lazyeval_0.2.0      NLP_0.1-11          mnormt_1.5-5
[31] mime_0.5            magrittr_1.5        nlme_3.1-131        SnowballC_0.5.1     foreign_0.8-69      tools_3.4.1
[37] LCFdata_2.0         munsell_0.4.3       cluster_2.0.6       bazar_0.1.6         kimisc_0.3          bindrcpp_0.2
[43] compiler_3.4.1      rlang_0.1           nloptr_1.0.4        grid_3.4.1          spam_2.1-1          htmlwidgets_0.9
[49] labeling_0.3        base64enc_0.1-3     gtable_0.2.0        reshape2_1.4.2      R6_2.2.1            knitr_1.16
[55] bindr_0.1           modeltools_0.2-21   rJava_0.9-8         stringi_1.1.5       parallel_3.4.1      Rcpp_0.12.12
[61] rgl_0.98.1          fields_9.0          rpart_4.1-11        acepack_1.4.1
```

Appendix H

Publications

H.1 Behavior Research Methods publication

Behav Res
DOI 10.3758/s13428-016-0750-0

A cluster-based approach to selecting representative stimuli from the International Affective Picture System (IAPS) database

Alexandra C. Constantinescu[1] · Maria Wolters[1] · Adam Moore[1] · Sarah E. MacPherson[1,2]

Abstract The International Affective Picture System (IAPS; Lang, Bradley, & Cuthbert, 2008) is a stimulus database that is frequently used to investigate various aspects of emotional processing. Despite its extensive use, selecting IAPS stimuli for a research project is not usually done according to an established strategy, but rather is tailored to individual studies. Here we propose a standard, replicable method for stimulus selection based on cluster analysis, which re-creates the group structure that is most likely to have produced the valence arousal, and dominance norms associated with the IAPS images. Our method includes screening the database for outliers, identifying a suitable clustering solution, and then extracting the desired number of stimuli on the basis of their level of certainty of belonging to the cluster they were assigned to. Our method preserves statistical power in studies by maximizing the likelihood that the stimuli belong to the cluster structure fitted to them, and by filtering stimuli according to their certainty of cluster membership. In addition, although our cluster-based method is illustrated using the IAPS, it can be extended to other stimulus databases.

Electronic supplementary material The online version of this article (doi:10.3758/s13428-016-0750-0) contains supplementary material, which is available to authorized users.

✉ Alexandra C. Constantinescu
caterina.constantinescu@ed.ac.uk

1 Department of Psychology, School of Philosophy, Psychology and Language Sciences, University of Edinburgh, 7 George Square, Edinburgh EH8 9JZ, UK

2 Centre for Cognitive Ageing and Cognitive Epidemiology, University of Edinburgh, Edinburgh, UK

Keywords IAPS · International affective picture system · Cluster analysis · Stimulus selection · Emotion

Introduction

It is now widely accepted that emotion plays a critical role in human psychology and is inextricably entwined with behavior and cognition. Yet, a major challenge that emotion researchers face is conceptualizing the relationship between various kinds of emotions and mapping their collective impact on other psychological processes (e.g., Ito, Cacioppo, & Lang, 1998; Lane et al., 1997; LeDoux, 1996). Perhaps the most widely used tool in this pursuit is the International Affective Picture System (IAPS; Lang, Bradley, & Cuthbert, 2008), which consists of 1,182 images and is designed for the experimental study of affective processing. It is based on the PAD model, involving pleasure/valence arousal, and dominance—a three-dimensional framework for measuring emotions (Mehrabian, 1996; Russell & Mehrabian, 1977). The validity of this theoretical model has accumulated a wealth of empirical evidence over time, and the number of citations for the database and instruction manual is now approaching 3,300, indicating a continued and robust research community surrounding it.

Using the IAPS database is particularly attractive due to the large variety of stimuli offered, as well as the chance to replicate and compare findings more easily between studies. Following the PAD model, each complete IAPS case is associated with normative (average) ratings for pleasure/valence (i.e., how positive or negative an image is), arousal (i.e., how alerting or calming an image is), and dominance (referring to the viewer's perceived amount of control in relation to the stimulus displayed). To exploit the flexibility offered by such a large number of stimuli, several typical approaches for image selection have been used, with some of the most common

Published online: 10 June 2016

being discussed below. However, it is important to note that most of these methods rely on assumptions about the underlying multidimensional structure of the database, and that violations of those assumptions can have profound consequences with respect to what inferences may be drawn from experiments using these stimuli. Specifically, if those assumptions are unsustainable, then some of the conclusions from the emotion literature may be questionable.

Establishing group cutoff points

This method consists of selecting cutoff values, which usually divide one of the three continuous PAD distributions into different categories. For instance, Mikels, Fredrickson, et al. (2005) distinguished between positive and negative stimuli on the basis of which IAPS images had valence ratings above or below 5, respectively, given the rating scale used to measure PAD dimensions in the IAPS contains nine points. Similarly, Xing and Isaacowitz (2006) considered the images with valence scores between 1 and 4 to be negative, those between 4 and 6 to be neutral, and those over 6 to be positive, with images very close to these cutoff points being excluded (Xing, personal communication, June 6, 2015).

A variant of using group cutoff points is selecting extreme groups of images. This consists of retaining the first *n* most negative/positive images (or an upper and lower group of images), as well as a group with minimal distances from what is considered a "neutral" score. For instance, one of the four types of emotion induction used in Zhang, Hui, and Barrett's (2014) study consisted of a combination of images and music, with some of the images being selected from the IAPS stimuli according to their rank (most positive, most negative, or most neutral).

Another extension of the cutoff point method was used by Lithari et al. (2010), who combined it with graphical presentation and selected images on the basis of how they were organized within a 2-D space. Four quadrants were formed through the crossing of the valence and arousal nine-point axes at a score of 5, and each quadrant was considered to represent a separate group of stimuli.

The cutoff point approach is best suited to research questions that focus on only one dimension of the PAD model. Although carefully chosen combinations of cutoff points may be adequate when a study focuses on only one or two dimensions, this strategy becomes unwieldy when researchers intend to systematically vary all three dimensions at the same time. Moreover, the use of cutoff points in this fashion tacitly assumes that the noncontrolled dimension(s) has (have) no effect on information processing or behavior that is relevant to the researchers' interests—an assumption that is risky at the best of times. Finally, another implicit assumption, for which there appears to be no clear evidence, is that the groups formed using the cutoff points can approximate the internal structure of the IAPS data correctly.

Discretization and crossing/controlling dimensions

This method refers to cutting the continuous PAD dimensions associated with the IAPS into *n* categories. Subsequently, within one such category, one may repeat the procedure on the basis of the remaining dimensions. For example, after cutting valence ratings into three categories, one may then attempt to find images of varying levels/categories of arousal within, for example, the most pleasant valence category. Alternatively, one may attempt to control one dimension within another—for example, finding one category with relatively constant arousal within the most pleasant valence category.

For instance, Tomaszczyk, Fernandes, and MacLeod (2008) chose IAPS stimuli on the basis of their valence ratings, but in addition attempted to cross different levels of arousal within the valence categories (see also Anderson, Siegel, & Barrett, 2011). Similarly, Aguilar de Arcos, Verdejo-García, Peralta-Ramírez, Sánchez-Barrera, and Pérez-García (2005) selected five categories of images for eliciting emotional experiences, including one neutral valence category with low arousal, and positive and negative valence categories, each with either a lower or a higher arousal level. Finally, Perri et al. (2014) divided the IAPS stimuli into positive, negative, and neutral categories based on their valence scores, with the first two of these categories presenting high levels of arousal. The neutral-valence pictures were selected to present low arousal.

If attempting to cross PAD dimensions in a factorial design in this manner, the assumption is made that the PAD dimensions are orthogonal (i.e., uncorrelated), which is not what the IAPS data suggest (Bradley & Lang, 2007). Similarly, attempting to control dimensions assumes that groups of images exist within the IAPS that vary in terms of one dimension, but not another. This is also generally not feasible, given that the correlated PAD dimensions tend to vary *together*. Finally, as is the case when using cutoff points, this method cannot easily accommodate the use of all three PAD dimensions simultaneously, usually resulting in dominance scores being ignored. Although it is correlated with the other two PAD dimensions, dominance represents a distinct entity within the model, and thus can itself account for some variation in affective ratings (Bradley & Lang, 1994). Therefore, if dominance scores are ignored, this variation would be excluded from the image selection process, which poses risks for its validity.

Content selection

This type of stimulus selection based on content is usually combined with one of the previously discussed methods. For instance, Bernat, Patrick, Benning, and Tellegen (2006) selected erotic and adventure scenes as pleasant, and violent or threatening images as unpleasant stimuli. Neutral images were chosen to portray common objects or inactive people, and so on. In addition, this strategy was combined with dimension discretization/crossing, leading to groupings of pleasant and unpleasant images with low, medium, or high arousal levels (see also Tomaszczyk et al., 2008). In another study, Hamann, Ely, Hoffman, and Kilts (2002) selected IAPS images on the basis of their content: Pleasant pictures were chosen to depict erotic scenes, food, or agreeable animals and children. Negative images were selected thematically to include mutilated bodies, violence, and so forth. In parallel, high-interest images included exotic parades and surrealistic scenes, and low-interest images included plants or household scenes.

In addition, Eizenman et al. (2003) emphasized the thematic selection of IAPS images: Four categories were selected to include images considered neutral, dysphoric, threatening, or socially themed. However, the authors also relied on valence ratings to guide their selection procedure, so that neutral images were selected to have valence scores close to 5, threatening/dysphoric images ranged between valence scores of 2 and 4, and the social themes presented a range between 6 and 8 on the same scale. They also aimed to control variations in arousal levels by allowing maximum differences of two points across the images in each of the four categories. The content selection method does not place strong assumptions on the data on its own; however, it is usually used conjointly with the content selection, discretization and crossing/controlling dimension methods, which do.

An alternative image selection method based on cluster analysis

The present work offers an alternative strategy for image selection based on clustering algorithms, which can be used with all three PAD dimensions simultaneously. To our knowledge, such algorithms have been used to categorize participant responses from individual studies (e.g., for classifying brain regions with differential response patterns to disgusting vs. neutral images—Deen, Pitskel, & Pelphrey, 2011; or for grouping participants in terms of their risk for alcohol abuse, on the basis of heart rate variability in response to IAPS emotional stimuli—Mun, von Eye, Bates, & Vaschillo, 2008), but not to group or select images on the basis of normative data.

In this article, we argue that clustering methods constitute a valuable means for creating experimental stimulus groups based on the IAPS normative data, by ensuring that group formation is optimized according to various measures (e.g., maximizing the distances between the different groups or the likelihood that cases belong to a certain group). This can boost the level of statistical power achieved in studies, since the larger the differences between levels of the treatment, the higher the chances of finding significantly meaningful effects (see Hallahan & Rosenthal, 1996, p. 495).

In addition to using more objective criteria for group formation, relative to entirely "manual" methods, clustering algorithms can also capture the particular structure of the IAPS data, and thus provide image classifications that are more empirically principled. This can allow experimenters to guard against confounds in the form of heterogeneous, systematically underpopulated, or "artificial" categories of stimuli, which cannot be adequately supported by the IAPS database. For instance, IAPS images are often divided into three groups based on valence. However, if this three-group structure is not an adequate fit for the IAPS normative data, images may be grouped inappropriately. Thus, if multiple types of negative material exist within the IAPS, creating only one category of negative images would risk blending these together, with unpredictable consequences for study results and the validity of any inferences based on them.

In addition, without consulting the structure of the IAPS data (which clustering methods are sensitive to), it might be tempting to resort to a factorial design combining three ordered levels of valence (low, neutral, and high) with as many levels of arousal. In this situation, it would be difficult to find enough images populating the intersection between low valence (i.e., negative images) and low arousal (i.e., relaxing images), due to the correlation between these two dimensions. Indeed, such a category could thus be deemed "artificial," as it would ignore the essential correlations between PAD dimensions.

Consequently, clustering methods can provide information on both the quantity and quality of stimulus categories that can realistically be supported by the structure of the IAPS normative data. Although such algorithms can be flexibly adapted to extract a predetermined number of groups, usually they are allowed to follow an exploratory strategy constrained by the overall structure of the data set. That is, they will find the "best" number of stimulus clusters/groups, subject to some optimization constraints. This is a point of departure from the typical selection methods discussed above, in which a top-down process is often used to identify three image categories fitting the notions of "negative," "neutral," or "positive." Finally, clustering algorithms can limit the amount of labor associated with stimulus selection, especially when research hypotheses involve more than one feature being taken into account at the same time (i.e., dominance, as well as valence and arousal). Indeed, by minimizing this difficulty, the method we propose below allows researchers to expand the scope and complexity of their hypotheses, and thus more easily test their theories.

Our hypothesis is that the IAPS data present a discernible, meaningful structure that can be capitalized upon by using cluster analysis to produce stimulus groups for experimental use. Here we tested several clustering approaches against one another, and propose a stepwise strategy for filtering and classifying IAPS images for subsequent experimental use. The family of clustering algorithms (or data-mining techniques) is extremely diverse and easily warrants entire books dedicated to them (for more detailed discussion, see Jain & Dubes, 1988; Kantardzic, 2011; and Kaufman & Rousseeuw, 2005). However, due to their widespread use and popularity, we focus on several approaches in particular. We will now briefly describe each of these approaches; readers interested in a more in-depth coverage may refer to the supplementary material.

The first approach is *k-means clustering*, which involves selecting k random seeds (i.e., random points in the space defined by the dimensions of the stimuli) and assigning the closest cases to them, leading to the formation of k groups. Afterward, the group mean (i.e., centroid) is computed, and cases are reassigned to groups on the basis of closeness to this value. This process will reiterate until the classification has settled into a stable solution (i.e., when the data points no longer change their memberships after the centroid computation). This is a hard partitioning method, meaning that all cases are included in their respective clusters with a probability of 1, and it does not provide a direct indication of the number of clusters existing in the data (Hartigan & Wong, 1979; MacQueen, 1967; Xu & Wunsch, 2009). Instead, various subsequent indices are used to suggest the number of clusters that would be appropriate for a given dataset. However, these do not take parsimony into account, and so may show little consistency or be prone to inflating the number of clusters. In order to establish clusters of images that could later be used as the levels of an "emotional content" independent variable, we tested k-means clustering because of its efficiency, simplicity, and wide use (Jain, 2010).

Another popular option is *hierarchical clustering*. This is an agglomerative method whereby individual cases begin by being designated as their own cluster (i.e., clusters of one data point each; Borcard, Gillet, & Legendre, 2011; Xu & Wunsch, 2009). Using one of multiple *linkage methods*, cases get merged progressively into ever-larger clusters, until all of the cases belong to just one, overarching cluster. Similarly to k-means, no indication is given about the suitable number of clusters in the data, so that with the aid of various statistical criteria (again not considering parsimony, and possibly conflicting in their recommendations), it is largely up to the researcher to decide where along this progression to stop and retain the corresponding number of clusters. Hierarchical clustering is also a hard clustering method, in which each case is assigned to one cluster exclusively, rather than being assigned a probability of membership.

A third option that is gaining in popularity is *model-based clustering*. This represents a form of hierarchical clustering that also involves an expectation-maximization (EM) procedure (for a primer on EM, see Do & Batzoglou, 2008). Unlike k-means, or hierarchical clustering per se, this is a soft clustering method, whereby cases are assigned to clusters with a certain probability (uncertainty) of membership. This can allow researchers to systematically control for the degree of typicality a stimulus exhibits in terms of the clustering dimensions used: A stimulus with higher uncertainty will be less representative of its cluster, and may introduce additional noise into experimental results. Also, in contrast with the two previous approaches, model-based clustering simultaneously provides both a clustering solution for the data and a straightforward method for determining the optimal number of clusters k. For this purpose, model-based clustering (implemented in the mclust R package: Fraley & Raftery, 2006) provides Bayesian Information Criterion (BIC) values and considers the optimal number of clusters for a given dataset to be whichever value of k maximizes[1] this criterion. Therefore, one of the distinctive features of this method is that it takes parsimony into account in the attempt to reduce the unnecessary inclusion of components (clusters) into the model.

To summarize, in this article we focus on three types of clustering—namely k-means, hierarchical, and model-based clustering—each of which differs in the approach taken to assigning case membership (and whether that membership is probabilistic or absolute). Moreover, the first two approaches do not intrinsically provide a clear criterion for determining the final number of clusters, and so admit a variety of methods for deciding this (see below and in the supplementary material). We tested each of these methods on the IAPS data in order to: (a) gain more insight into the internal structure of the database; (b) identify any common patterns in clustering solutions across the different algorithms; (c) select the most suitable algorithm of the three and retain its clustering solution, and lastly; (d) extract a fixed number of representative IAPS images from the final clustering solution for use in further experiments.

Subsequently, we employed various validation techniques, to select one clustering method as the most appropriate for the IAPS dataset. After selecting one such clustering algorithm, we extracted the best exemplars from each resulting cluster, which we then propose as the final selection of stimuli that researchers may wish to use in subsequent work.

Method

Dataset characteristics

In this study, we focused on the IAPS normative data gathered from both male and female participants, in which PAD ratings were collected using three (nonverbal) 9-point Likert scales

[1] The formula employed by Fraley and Raftery (2006) uses the negative of deviance, so that BIC here needs to be maximized rather than minimized, which is more common.

(using the Self-Assessment Manikin, or SAM; Bradley & Lang, 1994; Lang, 1980) and a sample of approximately 100 US students, depending upon the image. In our analysis, we included all three PAD dimensions that are available within the IAPS data, to create stimulus groups that account for the maximum amount of variance in participant responses (Bradley & Lang, 1994). Despite the large correlations between dominance and the other two PAD dimensions (see Fig. 1), dominance did not perfectly overlap with them (e.g., if $r > .9$) either empirically or theoretically, further justifying its inclusion in subsequent analyses.

Duplicates

We evaluated the univariate distributions available within the stimulus database, and identified 12 duplicate cases within the normative data (overall including $N = 1{,}194$ cases, but with $N = 1{,}182$ unique cases), each associated with different scores on the PAD model (see Table 1 for a listing). These images were likely normed twice, as part of different image sets (Lang et al., 2008). As a consequence, we replaced these duplicated pairs with a single entry containing the averaged valence arousal, and dominance across the duplicates.

Missing values

In terms of missing values, only the valence and arousal dimensions contained complete data. However, of the two dominance distributions ("Dom1" and "Dom2")[2] included in the database, depending on which SAM rating scale was used in the measurement (Lang et al., 2008), "Dom2" contained considerably more missing data than "Dom1." Thus, we retained only the "Dom1" scale for further use,[3] to benefit from its more complete data. We reduced the overall dataset accordingly, leading to a sample size of $N = 942$.

Results

Preliminary analyses

Outliers Given the variety of emotional material included within the IAPS database, we employed a form of outlier identification as an objective means to filter out images exceeding the emotional intensity of stimuli expected in daily life, which could prove overly stressful for participants.

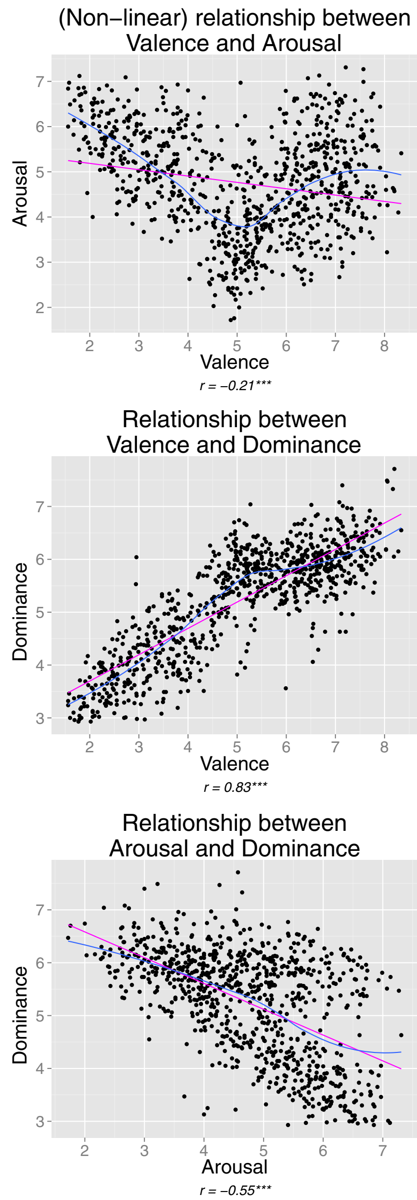


Fig. 1 Correlations between the pleasure/valence arousal, and dominance dimensions, with deviations from linearity that give rise to the specific shapes of the relationships

Outliers might also distort the clustering solutions (e.g., for *k*-means and model-based approaches), thus constituting an additional reason to identify and remove them. Specifically, outliers used with the model-based clustering might lead to a different number of clusters and/or alter the cluster memberships, without necessarily nesting outliers into a cluster of

[2] "Dom1" refers to the "classic" SAM dominance scale, whereas "Dom2" refers to a version of the dominance scale on which the SAM icon with the highest control presented a more assertive and dominant facial expression/posture than in the classic version.

[3] The correlation between "Dom1" and "Dom2" was remarkably high ($r = .98$), for the $N = 60$ cases measured on both versions of the dominance scale. Thus, we were able to safely use only "Dom1" in our analyses.

Table 1 IAPS duplicates and their valence arousal, and dominance ratings (devised using the Stargazer R package; Hlavac, 2013)

Description	Image Code	Valence	Arousal	Dominance	Set
Spider	1230	4.09	4.85	4.58	1
Spider	1230	4.61	4.03	5.60	2
Horse	1590	7.18	4.74	5.54	2
Horse	1590	7.24	4.80	5.62	3
Rabbit	1610	7.82	3.08	6.77	1
Rabbit	1610	7.69	3.98	6.52	2
Coyote	1640	6.27	5.13	5.22	1
Coyote	1640	6.16	5.18	4.91	2
Cow	1670	6.81	3.05	6.53	1
Cow	1670	5.82	3.33	5.63	2
NeutFace	2210	4.38	3.56	5.03	1
NeutFace	2210	4.70	3.08	5.23	2
Mutilation	3000	1.45	7.26	2.99	1
Mutilation	3000	1.59	7.34	2.73	4
Mutilation	3010	1.71	7.16	2.88	2
Mutilation	3010	1.79	7.26	2.88	3
EroticFemale	4220	8.02	7.17	5.33	2
EroticFemale	4220	6.60	5.18	5.90	3
EroticMale	4520	7.04	5.48	5.48	2
EroticMale	4520	6.16	4.80	5.73	3
AimedGun	6200	2.71	6.21	3.35	1
AimedGun	6200	3.20	5.82	3.49	2
Exhaust	9090	3.56	3.97	4.51	2
Exhaust	9090	3.69	4.80	4.72	3

their own (Fraley & Raftery, 2002; Hautamäki, Cherednichenko, Kärkkäinen, Kinnunen, & Fränti, 2005; Wu, 2012; Xu & Wunsch, 2009).

Using the R language (R Development Core Team, 2015),[4] all three univariate distributions were found to be nonnormal according to the Shapiro–Wilk test, so any method of determining outliers that was based on averages would probably be inappropriate (since the averages would not adequately represent the distribution). Hence, we opted for a more robust indicator: the Median Absolute Deviation (MAD; Leys et al., 2013).[5] Therefore, images that were more than 2.5 MADs away from the median, in either direction, were removed before further analyses could be conducted. No outliers could be identified using this method in the valence or arousal distributions, but interestingly, 32 images[6] were flagged as outliers due to their dominance scores, and were thus removed. This was done to avoid distorting the clustering solutions subsequently, and also to filter out potentially harmful material, in an empirically principled, replicable manner.

Representativeness/precision of measures Additionally, we implemented a measure to ensure the precision of the stimuli to be used: building 95 % confidence intervals (CIs) around the normative image ratings, to give an indication of how precisely the population means could be estimated, on the basis of the sample averages from the approximately 100 participants rating each image. We selected stimuli with CIs spanning no more than one point in total around the normative rating, which we considered to be sufficiently narrow, given that the three dimensions were measured on 9-point Likert scales. Using this criterion, 61 cases that were judged too imprecise were removed, since they could subsequently affect the inferences in our study; 46 images were removed due to the width of their CI on one dimension, 13 due to their CI width on two dimensions, and finally, two cases with CIs too wide on all three PAD dimensions simultaneously. After we had removed cases on the basis of both outlying values and CI widths, the sample size was reduced to $N = 849$.

Clustering techniques

After employing the filtration methods described above, three clustering procedures—*k*-means, hierarchical, and model-based clustering—were used to produce a set of coherent clusters that could be used in later primary research. For the reasons explained previously, the clusters were built on the basis of the normative ratings for all three available measures associated with the IAPS: valence arousal, and dominance.

***K*-means clustering** When using this method, various indices were consulted to identify what the appropriate number of clusters (k) should be, including the Caliński–Harabasz Index (Caliński & Harabasz, 1974), the Ball Index (Ball & Hall, 1965), and the Hartigan Index (Hartigan, 1975), which are all based on within-/between-cluster sums-of-squares calculations (i.e., minimizing the former and/or maximizing the latter to ensure cluster compactness and/or the separation between clusters), as well as the Simple Structure Index (SSI; Dimitriadou, Dolničar, & Weingessel, 2002; Dolnicar, Grabler, & Mazanec, 1999), and others. The general trends shown by some of these indices are presented in Fig. 2, where the nature of the dataset is such that various clustering indices detect different characteristics of the data and do not converge on any simple answer as to the "correct" number of clusters that should be extracted. For further details on these and other indices, please see the supplementary material.

[4] The R code for our analysis is available at www.github.com/CaterinaC/IAPSClustering2016.

[5] According to this method, acceptable values should lie between the median ± (x * MAD), where we opted for $x = 2.5$.

[6] IAPS codes: 3000, 3001, 3010, 3015, 3053, 3059, 3063, 3064, 3080, 3102, 3131, 3170, 3266, 3500, 3530, 6230, 6231, 6250, 6250.1, 6260, 6263, 6300, 6350, 6510, 6520, 9075, 9252, 9410, 9413, 9600, 9908, and 9940

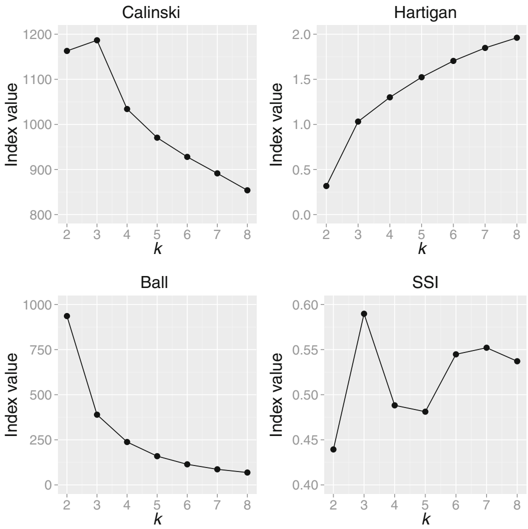


Fig. 2 Various clustering indices indicate different "optimal" values for *k*. These graphs may change slightly with every run of the clustering algorithm, due to the random seeds that *k*-means uses. As such, 100,000 repetitions were run on the *k*-means clustering algorithm each time, with a range for *k* from 2 to 8, and with the values of the Caliński, Ball, Hartigan, and SSI criteria computed each time (with the Ball criterion having to be minimized, unlike the other three criteria, which must be maximized). The average values for these criteria were then computed across all of the repetitions and indicated (left to right, and top to bottom) that three, eight, eight, and three clusters should be extracted, respectively

On the one hand, it may seem surprising that a subset of over 800 IAPS images may have several *k*-means clustering criteria peak for the number of only two[7] or three clusters, considering the amount of variation in both the content and scores of the IAPS images. However, this could be accounted for theoretically by the emergence of a dichotomous "Positive and Negative Affect" structure (PA/NA, developed more in the Discussion), sometimes accompanied by the natural emergence of an additional neutral cluster. In Fig. 3, both clustering solutions are displayed using color coding for each cluster in the 3-D space, and are shown to cover extensive areas of the 3-D space.

On the other hand, higher values for *k* might be more suitable for the data, as is suggested in Fig. 4, which shows that as the number of clusters increases, so does the amount of explained dissimilarity between the cases (calculated as 1 – *unexplained dissimilarity* or 1 – *within-cluster dissimilarity*). Thus, as the number of clusters increases, within-cluster homogeneity also increases. However, *k*-means does not penalize for the increasing number of clusters (unlike model-based

[7] Please refer to the supplementary material for more details on the measures of Connectivity and Average Silhouette Width that suggested this value.

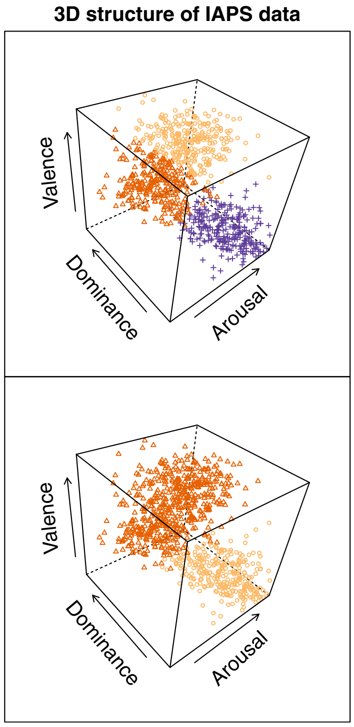


Fig. 3 Data structure of the IAPS images. It is worth noting that large portions of the 3-D space remain unpopulated, signaling either that the IAPS does not cover those combinations between valence arousal, and dominance, or that photographic material in general would have difficulty with this

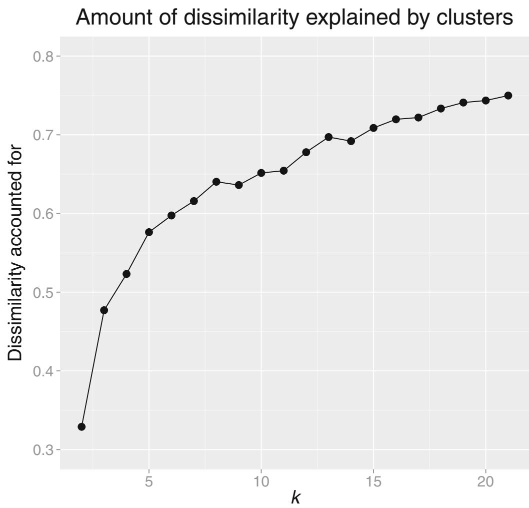


Fig. 4 The amount of dissimilarity (as computed using the R package clue: Hornik, 2005) between cases is accounted for by ever-increasing values for *k*

clustering), so that, conceivably, the total amount of dissimilarity would only be explained when the number of clusters equaled the number of cases. In other words, there is no single, definitive cutoff to determine which value of k best fits the data.

Since there may be arguments against using either a very small (e.g., $k = 2$ or even $k = 3$, with too many heterogeneous cases blended in the same group, as shown in Fig. 3) or a very large number of emotional categories (e.g., $k \geq 8$, leading to a very fragmented and unparsimonious structure, with relatively few cases per cluster), we now turn to the other clustering methods for additional solutions.

Hierarchical clustering Jointly testing various linkage methods (i.e., strategies for progressively merging clusters, described in more detail in the supplementary material) and distance metrics allowed us to find the combination yielding the clustering solution with the highest degree of similarity to the original data (or matrix containing the distances between every pair of IAPS cases). We found that Average Linkage (i.e., merging clusters based on the average distance between their points) paired with correlation-based distances (i.e., assigning cases to clusters on the basis of correlations) produced the results most similar to the original distance matrix (cophenetic correlation, $r = .91$). Consequently, this combination was the most suitable for the IAPS data, and shows how essential PAD relationships are when determining how to group the IAPS images. The next best result was attained by Single Linkage (in which cluster-merging depends on the distance between the closest points belonging to different clusters), again combined with correlation distances ($r = .87$). Thus, after having reconfirmed the importance of the PAD correlations and identified the most suitable hierarchical agglomeration method for this dataset, we proceeded to determine the most appropriate number of clusters in the data.

In terms of connectivity, average silhouette widths, and Mantel optimality (briefly described within the supplementary material), a number of two clusters was suggested, whereas the Dunn Index indicated three. This corroborates the findings from some of the k-means indicators, and suggests the overall strength of the PA/NA structure within the IAPS, with or without an additional neutral cluster. However, as with k-means, some variability was to be found; for example, when using the elbow method for partitioning variance into clusters (using the GMD R package; Zhao & Sandelin, 2012), the optimal number of clusters (also based on average linkage) indicated was seven. Other clustering indices suggested nine clusters; however, still others provided more discrepant results, indicating numbers ranging from four to 15, or as many as 30 clusters. Overall, the most endorsed options were two (perhaps three), or nine clusters. For more information, please see the supplementary material.

Model-based clustering Model-based clustering yielded a mixture model containing five clusters of varying Volumes, Equal (ellipsoidal) shapes, and Varying orientations (VEV). This model/configuration was optimal in terms of BIC values: BIC = –6,341.11, relative to the global minimum BIC value[8] for other cluster numbers and configurations, BIC = –8,671.93 (for one spherical cluster, with either equal or variable volume, and the configurations abbreviated as EII and VII, respectively). The second best BIC value achieved was –6,343.72, for a VEV model with four components (clusters). Full details regarding the BIC values for all the models considered can be found in the supplementary material.

The five-cluster solution proposed by the algorithm is described in Table 2, in terms of cluster centroids, sample sizes, mixing proportions (i.e., proportion of the mixture/overall sample that has been assigned to each cluster), and average uncertainties. By-cluster boxplots are also displayed in Fig. 5, comparing the relative spreads of the clusters' valence arousal, and dominance univariate distributions. In addition, given the cluster centroids presented in Table 2, it is apparent that this clustering solution presents a symmetrical format: two negative clusters (one more so than the other), one neutral cluster, and two positive clusters (one more so than the other).

Finally, we assessed whether the assumption of multivariate normality held for these clusters, and found that, overall, the clusters presented ellipsoid shapes consistent with this assumption, with some further evidence also added by various multivariate normality tests. Please see the supplementary material for details on testing the assumptions required for model-based clustering.

Validating the clustering solutions

After having employed three candidate methods—k-means, hierarchical, and model-based clustering—we proceeded to compare them on the basis of various validation techniques (full details are in the supplementary material), to select just one for further use. Given that variations were observed in terms of the "optimal" number of clusters suggested for k-means and hierarchical clustering by each clustering index, we deemed it appropriate to emphasize and pursue model-based clustering, which proved less affected by these issues, and also provided more information about the classification in the form of membership uncertainties. For a more meaningful comparison between the methods, parsimonious clustering solutions were formed using each of the three algorithms for a number of $k = 5$ clusters, as was suggested by model-based clustering.

Finding a stable structure within the data, across methods

Assuming that the IAPS data present a clear, discernible structure, all of the clustering algorithms should in principle be able

[8] Mclust() seeks to *maximize* BIC values, given that it uses the negative of deviance.

Table 2 IAPS cluster centroids, cluster sample sizes, mixing proportions (or the proportion of total cases assigned to each cluster), and average uncertainties, extracted using model-based clustering

Cluster	Valence	Arousal	Dominance	*N*	Mixing Prop.	Average Uncertainty
1	3.56	5.18	4.34	244	.29	.09
2	7.27	4.69	5.96	71	.08	.26
3	2.27	5.87	3.55	71	.08	.24
4	5.05	3.31	5.84	152	.18	.17
5	6.44	4.82	5.90	311	.37	.14

to identify this structure despite their computational differences. To check this, we assessed the extent to which model-based clustering yields membership assignments that overlap with those from the other two competing methods.

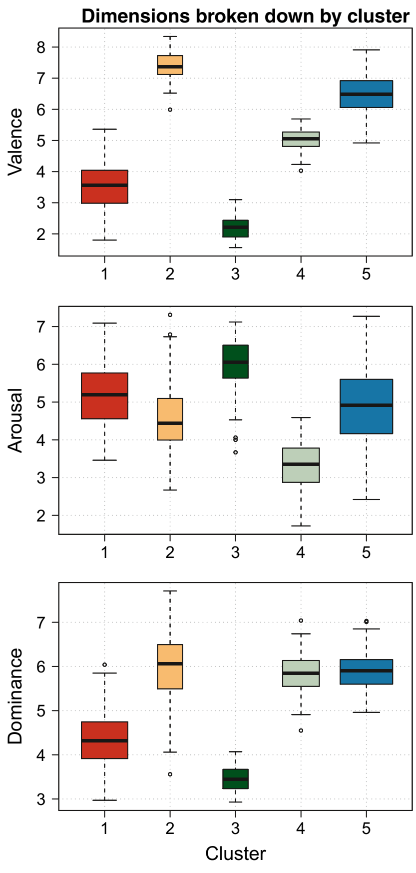


Fig. 5 Cluster boxplots, for each dimension. The boxplots indicate, for each cluster (coded by colors), the spread of cases assigned to it, in terms of valence arousal, and dominance. The boxplot widths are proportional to the cluster sample sizes

The Variation of Information criterion (VI; Meilă, 2007) suggests that not much information is to be gained/lost when moving from one classification to another (i.e., there is considerable similarity between partitions of five clusters, regardless of the algorithm used to produce them), with the normalized VI between model-based and *k*-means clustering = .176 and the VI between model-based and hierarchical clustering = .217 (please see the supplementary material for details). This finding was corroborated by the relatively strong association found between partitions using Cramer's ϕ (between the *k*-means and model-based classifications, ϕ = .704, and between the model-based and hierarchical classifications, ϕ = .516). Therefore, on the basis of the VI and Cramer's ϕ, there is considerable similarity between the five-cluster solutions provided by the different algorithms. However, for further results, including those based on the Adjusted Rand Index (ARI; Hubert & Arabie, 1985), please refer to the supplementary material. Thus, on the whole these results constitute moderate evidence that a specific data structure can be identified in the IAPS, given the level of agreement between the clustering methods.

Evaluating the stability of the model-based clustering solution We assessed the stability of the clustering solutions using various criteria, including split-half validation (i.e., dividing the IAPS data into two random halves and computing the level of association between the partitions created independently on these halves of the data) and jackknife validation (i.e., removing 10 % of the IAPS data randomly across a few thousand repetitions and assessing changes in the structure of the clustering solutions). Overall, model-based clustering performed well, with a high degree of association present between how the random halves of the data were clustered, suggesting that the stimulus groups identified were well-supported. In terms of stability after the random removal of 10 % of the data points, model-based clustering also outperformed both *k*-means and hierarchical clustering, for which typically only one cluster was then identifiable in the data (i.e., no grouping of the data points could be achieved after the removal of data points using these methods). For more details on these and further analyses, please refer to the supplementary material.

Selecting equal numbers of cases from each cluster

Given that the five clusters provided by model-based clustering differed in size, a procedure was required to sample equal numbers of cases from each cluster that would represent their respective cluster to the highest degree. Since levels of certainty are also provided for each image during the model-based clustering process, these could be used to create a hierarchy in terms of how likely it was for each image to belong to the cluster it was assigned to.

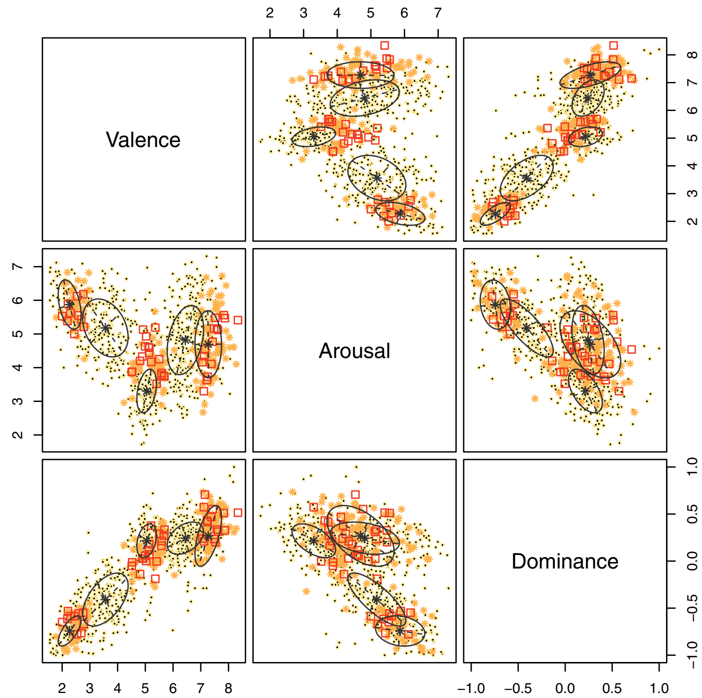


Fig. 6 Bivariate scatterplots showing the default classifications of cases and the uncertainties provided by Mclust() in R. The uncertainties are coded using one of three symbols: ringed black dots for candidates with a high certainty of cluster membership; orange (light gray) asterisks for less clear cluster memberships; and red (dark gray) squares for cases to avoid using as stimuli, with very unclear memberships. Point size is an additional indicator for the level of classification uncertainty, with larger points indicating higher uncertainty

Consequently, a given number of images could be selected according to their rank in this hierarchy (i.e., the first *n* most likely cluster members). Figure 6 shows the default distinction made by Mclust(): Cases with uncertainties below the 75th percentile are considered acceptable, uncertainties between the 75th and 95th percentiles are risky candidates, and those over the 95th percentile should not be used, as they do not show clear membership to a given cluster. We made the same distinction in our final results, available online for download in the repository at www.github.com/CaterinaC/IAPSClustering2016, where we indicate which IAPS images were assigned to which cluster, as well as the level of uncertainty associated with this classification—particularly, which uncertainties were above or below the 75th percentile (i.e., whether or not they should be sampled for research). These results are suitable for researchers to use in most research contexts.

In our example, only the first 20 cases in the hierarchy of uncertainties were retained for closer inspection. These can be judged as the best representatives for each given cluster, and are portrayed in Fig. 7, with the first five of each cluster also displayed in Fig. 8, where they are shown to be meaningfully related to one another.

A comparison between our method and ad-hoc approaches to selecting IAPS stimuli

Studies relying on more typical, ad-hoc methods for sampling IAPS stimuli may face several risks. On the basis of a Google Scholar search for "IAPS images," we selected a small

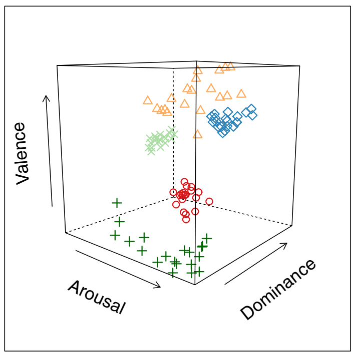


Fig. 7 Selection of the 20 most likely IAPS cases per cluster, for the $k = 5$ clustering solution. The color coding was chosen to be consistent with Fig. 8 below

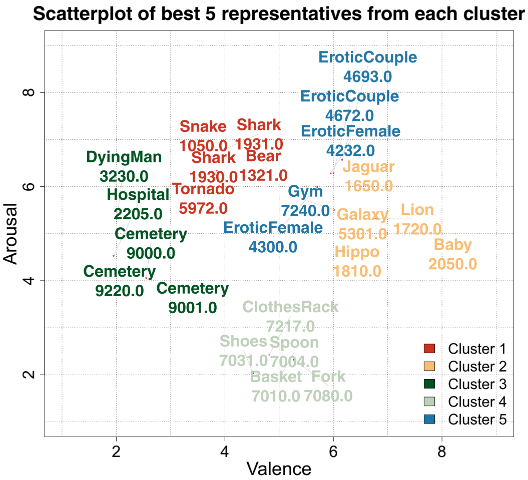


Fig. 8 Selection of the five most likely IAPS cases per cluster, for the k = 5 clustering solution, along with IAPS image codes. The color coding was chosen to be consistent with Fig. 7 above

number of studies randomly from several pages of results. However, we only retained articles that also specified the IAPS image codes used, rather than simply the average PAD values for the images selected. We then assessed how the categories used in these studies matched our own.

First, as is shown in Table 3, the images intended to represent different affective categories in these studies sometimes share the same clusters that our model-based clustering uncovered. For instance, in the Glenn, Blumenthal, Klonsky, and Hajcak (2011) study, four images considered neutral and ten images considered pleasant all belong to one of our *positive* clusters (i.e., Cluster 5; see also Table 2 for cluster descriptions).

Second, the negative or positive stimulus groups used in studies tend to pool together stimuli that our method has distinguished as reflecting two types of positive or of negative material. For example, the Koenigsberg et al. (2010) study used a group of stimuli wholly considered to be negative; however, our method divided these between two separate clusters—one that is mildly negative and moderately arousing, and one that is more negative and more arousing, and with lower dominance than the former cluster.

In some cases, a single stimulus category (i.e., neutral, on the basis of the research reviewed in Table 3) may spread across three or four of our clusters. For instance, in the study by Most et al. (2005), the neutral category in fact included eight mildly negative images, 27 neutral images, and 20

Table 3 Stimulus groups used in various studies, redistributed according to our method

Study	Stimulus Categories Used	Total No. Stimuli Used per Category	Image Codes Unaccounted For	Stimulus Redistribution According to Our Method, According To:							
				Missing Dominance Score Excluded	Outliers Excluded	Wide CI Excluded	C1 (−)	C2 (++)	C3 (− −)	C4 (±)	C5 (+)
Glenn et al., 2011	Pleasant	18	–	–	–	3	–	5	–	–	10
	Neutral	18	–	1	–	–	–	–	–	13	4
	Unpleasant	18	–	–	6	3	4	–	5	–	–
Mikels, Larkin, et al., 2005	Pleasant	80	17	–	–	10	–	22	–	–	31
	Neutral	80	8	–	–	–	7	5	–	41	19
	Unpleasant	80	–	–	3	2	71	–	4	–	–
Most et al., 2005	Neutral	55	–	–	–	–	8	–	–	27	20
	Negative	39	–	–	11	6	5	–	17	–	–
Stins & Beek, 2007	Neutral faces	14	–	1	–	–	2	–	–	10	1
	Neutral household items	15	–	1	–	–	–	–	–	14	–
	Erotic	17	–	1	–	7	–	–	–	–	9
	Family scenes	12	–	1	–	2	–	8	–	–	1
	Mutilation	11	–	–	6	–	–	–	5	–	–
	Fear	18	–	1	5	1	9	–	2	–	–
Koenigsberg et al., 2010	Negative	47	–	6	7	4	13	–	17	–	–
	Neutral	49	–	10	–	–	4	–	–	25	10

The clusters from one through five are represented by C1, C2, C3, C4, and C5, and refer to those described in Table 2. For the same columns, between parentheses we have included concise information about the valence of our clusters, ranging from very positive (++) to very negative (− −). This has been done merely to aid interpretation of this table; however, as we previously stated, all three of the PAD dimensions were important in determining how the IAPS stimuli were assigned to these clusters, not only valence.

mildly positive images, according to our method. Another example is the study by Mikels, Larkin, Reuter-Lorenz, and Carstensen (2005), in which a category of neutral images intended to differ only in brightness actually belonged to four different emotional clusters within our own classification.

In addition, from Table 3, it is also apparent that without filtering images on the basis of 95 % CIs, less reliable image stimuli can be included in studies. For instance, in the case of the Stins and Beek (2007) study, seven less reliable (in terms of confidence interval widths) images were included in the group of erotic stimuli. Similarly, images have also been selected without taking into consideration dominance—including some images we excluded precisely because their norms for dominance were missing. Finally, IAPS data outliers have also been included in studies, which could pose some ethical risks, due to their emotional intensity, and warrant closer inspection.

Discussion

A variety of research areas rely on stimulus databases for experimental use. The IAPS is one such widely used database, having currently amassed approximately 3300 citations in Google Scholar (April, 2016). Yet, despite its extensive use, a standard stimulus selection strategy from the IAPS has yet to be devised—one that can easily take into account all three PAD dimensions simultaneously, and provide a stimulus grouping that is both empirically principled and optimal in terms of various statistical measures.

In this article, we proposed such a method based on the following sequence of steps: filtering out stimuli that constitute outliers or duplicates, and those with CIs wider than a preset criterion; creating stimulus categories using different clustering algorithms; and finally, validating these categories against several measures. Within the procedure we propose, we placed special emphasis on model-based clustering, an inferential method that provides not only a classification of the stimuli, but also an uncertainty estimate for each stimulus assigned to a cluster. Examining these uncertainty estimates allows researchers to control for how well stimuli reflect their underlying category and to select only those stimuli that reflect their cluster in the most meaningful way.

Filtering out stimuli prior to clustering

As a first step toward creating a selection of stimuli for experimental use, the MAD has proved to be a useful tool for identifying stimuli that may be ethically questionable, due to their violent or threatening nature. In addition, Grühn and Scheibe (2008) found that IAPS ratings for negative images tend to get more extreme with age. Thus, as a precautionary measure, filtering out outliers using the MAD might have to be considered more carefully depending on what sample/population the stimuli are aimed at, as the same IAPS image might be more distressing for one category of participants than another.

Using the MAD, we were able to exclude 32 images due to their particularly low dominance scores (i.e., in the case of highly violent images, with an average valence level of 1.98—e.g., image 3001, a headless body; 3131, mutilation; 3170, a baby with a tumor; etc.). Interestingly, these same cases were not flagged as outliers given their scores on the other dimensions. This provides further evidence that dominance scores reflect a different process of emotional evaluation and should be considered more frequently when selecting IAPS images. Relatedly, dominance is believed to be more easily distinguishable from the other two dimensions in social situations (rather than photographic material; Bradley & Lang, 2007, p. 32), further supporting its general inclusion in stimulus selection procedures, as an additional contributor to emotional experiences.

The large standard deviations associated with the ratings for most stimuli from the IAPS have usually resulted in wide 95 % CIs (spanning more than one point on the nine-point Likert scale used for ratings). However, within our overall approach based on CIs, other (more or less conservative) criteria may also be applied regarding the width of these CIs, depending on researchers' specific aims. This type of verification has proven to be highly useful either for deciding which stimuli to retain for the subsequent clustering procedure, and for a better appreciation of the amount of variability in the individual IAPS ratings leading to the normed means. Although we are unable to give an exact reason why some of the stimulus norms were insufficiently precise, on the basis of our criterion, these results clearly suggest a verification as simple as this should become a more standard practice when selecting stimuli from stimulus databases.

We would stress that it is possible for any emotional stimuli database to present these same concerns. This is because emotional stimuli are conceivably very subjective, thus leading to the large standard deviations observed, and implicitly, the lower degree of certainty as to how they may be perceived by individual participants (e.g., image "EroticFemale" 4210 registered the highest standard deviation of all IAPS images, suggesting that reactions to it varied considerably). On the other hand, it is also possible these characteristics might be specifically related to the features of IAPS, but not of other emotional stimuli collections; thus, image quality and historic context, ecological validity, and so forth, may also be involved. Future work will be necessary to address this research question.

Clustering the stimuli

When using k-means and hierarchical clustering to classify IAPS images, the repartition of cases between clusters represents a separate step from choosing the "appropriate" number of clusters existing within the data. Our analysis showed that it is difficult to discern a clear cluster structure within the IAPS data. For example, in the case of *k*-means, the optimal value for *k* oscillated between two, three, or eight, depending on the clustering index used, and on the total number of clusters tested. Similarly, for hierarchical clustering, a number of two, three, seven, or nine clusters was indicated as suitable for the IAPS data, also depending on the index and number of clusters. It may seem surprising that a number of clusters as low as two, or even three, could be suggested by both *k*-means and hierarchical clustering, for a sample size as large as N = 849 images, varying considerably in terms of valence arousal, and dominance scores. However, the emergence of these solutions is understandable, for theoretical reasons and/or due to the shape of the IAPS data.

First, the k = 2 solution carries theoretical significance by corroborating principles used in the construction of the Positive and Negative Affect Schedule (PANAS; Watson et al., 1988), since the two emerging clusters can be interpreted as matching the Positive and Negative Affect components of the scale, which measure the corresponding affective moods with adequate reliability and validity. This similarity directly indicates that clustering methods can provide meaningful results, which can be validated against current practices and/or theory.

Second, the nonlinear ("U" shaped) relationship between valence and arousal can easily be split into three sectors, a characteristic that carries over into 3-D space, when dominance is added. Thus, one cluster is negative with higher arousal, another is neutral with lower arousal, and the third is positive, again with higher arousal. Although this three-cluster solution may appear similar to those from typical image selection practices (cutoff points and/or factorial designs, centered on selecting three valence groups: negative, neutral, and positive), it differs from these approaches in that it accommodates all three PAD dimensions simultaneously with ease, and also takes the structure of the data into account, without imposing unsustainable assumptions (i.e., independence of the PAD dimensions). In fact, even if hierarchical clustering did not provide the final classification of the IAPS data, it did reveal most clearly the importance of the PAD relationships, since using correlation-based distances always yielded the highest correlations with the original data for this clustering method. This suggests that the PAD correlations should always be taken into account when selecting stimuli from the IAPS, whereas using factorial designs without concern for them may simply lead to inappropriate groupings of stimuli, and subsequent experimental results that are difficult to interpret.

However, both of these solutions (k = 2 and k = 3) focus on the creation of just a few, large clusters, which would thus cover considerable portions of the 3-D affective space within the PAD model. As such, one large negative cluster would, for instance, include images with both moderate and higher arousal, or both moderate and lower dominance—leading to a lower degree of experimental control.

On the other hand, from a practical standpoint, the larger numbers of clusters (seven, eight, or nine) indicated by *k*-means and hierarchical clustering may be as intractable as the lower numbers, but for different reasons. Rather than blending together too many heterogeneous cases, when using a larger number of small clusters—the more clusters are extracted, the closer their centroids necessarily become and thus their "best representatives" are also drawn nearer. This can result in a potential reduction in statistical power. Also, more clusters (or treatment levels) would generally signify longer testing times and study expenses, which is not always feasible. Finally, smaller cluster sizes would be less useful for experiments requiring larger numbers of stimuli of the same type (i.e., from the same cluster).

In contrast to the previous two methods, model-based clustering uses a soft clustering approach, which provides an estimate for the degree of cluster membership (uncertainty) associated with each image. This allows for finer-grained control over stimuli used in experiments, which in turn can help make research inferences stronger. This method also provides additional flexibility in terms of adaptively distinguishing a variety of cluster configurations, thus being capable of a closer fit to the original data. In contrast, *k*-means would, for instance, favor spherical clusters in particular (Jain, 2010). Finally, unlike for *k*-means or hierarchical clustering, the optimal number of clusters in model-based clustering is assessed using the BIC, which penalizes for large numbers of clusters, and simplifies the process of choosing which number of clusters to extract from the data.

In our case, a number of five clusters was suggested, which also represents a good compromise from a practical standpoint. In addition, the clusters were determined to be of Varying volumes, Equal shapes (i.e., ellipsoidal, rather than spherical), and Varying orientations within the 3-D space. The cluster centroids also suggest that for participants, "neutral" images present medium levels only on the valence scale, rather than in the whole PAD model, as might have been assumed. Thus, neutral IAPS images tend to be somewhat lower in arousal and higher in dominance: For instance, a picture of a mug (IAPS code 7035) intuitively seems "neutral," but this translates into medium values only on the valence dimension (norm = 4.98), whereas the lower arousal (norm = 2.66) suggests a more calming influence, and the higher dominance (norm = 6.39) suggests very unchallenging content.

Equally, we have shown that two forms of negative and positive material exist, rather than one of each, which is the typical grouping used in research. For instance, we found that very negative content (e.g., "Mutilation", IAPS code 3030) presents very low valence (as expected) but, uniquely, higher arousal and lower dominance. Thus, collectively, these three components (and not just valence) seem to form what is usually perceived as "very negative" content. A second, milder, type of negative content was identified, as well, which still presents valence values below the scale midpoint, but less extreme arousal and dominance values (e.g., "Cigarettes", IAPS code 9832). Similarly, positive content can also be divided into two subtypes using our method: positive, more arousing content (e.g., "Erotic Couple", IAPS code 4693) and very positive, more serene/less arousing content (e.g., "Nature", IAPS code 5220)—with both of these categories being fairly similar in their mean-level dominance.

This five-cluster option generally benefits from empirical support based on the methods we employed to verify this. We first noted a moderate overlap between how the images were classified into five groups by *k*-means, hierarchical, and/or model-based clustering, depending on the measure used to assess the overlap. Although no structure is unanimously accepted within the IAPS data, measures such as the Variation of Information (VI) or Cramer's ϕ both suggested that $k = 5$ is relatively well-supported, even if each clustering method can shed its own perspective on the data (i.e., the amount of overlap was not maximal, which we discuss in more detail in the supplementary material).

Subsequently, to ensure that model-based clustering is indeed the most suitable algorithm for use with the IAPS data, we removed 10 % of cases randomly across a few thousand repetitions (using jack-knife validation), each time assessing how the optimal number of clusters changed. Ideally, if a robust clustering solution was found using a certain clustering algorithm, the removal of 10 % of the values should make little difference. In the case of *k*-means and hierarchical clustering, this frequently resulted in only one all-encompassing cluster being identified in the data, which was deemed inappropriate. In contrast, model-based clustering showed more stability, and most often suggested $k = 3$ (followed by $k = 4$) as the optimal solution in this case. However, cross-tabulations showed that these model-based solutions were very closely correlated to the $k = 5$ solution achieved on the full dataset, and did not present any deeply concerning changes such as the cluster structure collapsing entirely (i.e., when we found just one cluster using the other two methods). Therefore, the differences seen in the values of *k* most likely reflect the fact that one or two clusters from the $k = 5$ solution were collapsed due to the induced data attrition (–10 %), but that similarities between the solutions nevertheless remained robust.

Finally, when predicting the clustering structure of a random 50 % of values based on that of the other 50 % (using split-half validation), and comparing this prediction to the observed model-based classification of the target half, the two matched very closely. On the basis of all these indicators, we concluded that the five-cluster mixture model is well-supported by the IAPS data.

Method summary and recommendations for use

As an outline for our method, we recommend first inspecting the IAPS images and filtering out duplicates, outliers, and images with CIs larger than a preset criterion (we opted for one point in total, on the Likert scales used for the IAPS norms, but researchers may be more conservative if they have specific reasons for this). Subsequently, on the basis of the findings detailed above and in the supplementary material, we recommend resorting to a model-based clustering algorithm, which will nest the remaining images into five clusters, while also taking into account arousal and dominance in the creation of these clusters, even if researchers may only be explicitly interested in, for instance, valence.

Regarding any more practical issues that may arise, we recommend maintaining this well-supported, five-cluster structure even if researchers may be interested in comparing fewer categories. For instance, assuming that a study is aiming to compare the effects of positive versus negative valence on an outcome variable, just two of the five clusters may be used, which are farthest apart on this dimension, rather than altering the clustering solution to provide just two clusters in total.

Given that model-based clustering is a soft clustering method, cases were also assigned a level of certainty for belonging to their cluster. Unequal cluster sizes (with some of them being perhaps too large to be used in an experiment in their entirety) led to cases being sorted in descending order of their certainty of membership. This enabled us to select a constant number of images per cluster for subsequent use in an experiment—those at the top of the hierarchy formed (i.e., with the highest certainty of membership, or equivalently, with the lowest uncertainty). Besides providing the ability to flexibly tailor this constant to the requirements of individual studies, these stimuli can also act as the best representatives of their respective clusters.

For illustrative purposes, five to 20 cases per cluster were sampled in the order of their certainty of belonging to their given cluster. This resulted in groups that are intuitively meaningful, with one very negative cluster including death-related scenes (e.g., hospital, cemetery, dying man); a second negative cluster including dangerous agents, which was higher in dominance than the former one (e.g., snake, bear, shark); one neutral cluster that was low in arousal and higher in dominance (e.g., spoon, shoes, basket); one positive cluster including arousing scenes (e.g., erotic scenes, gym); and finally, another very positive cluster including less arousing "natural" scenes (e.g., hippo, jaguar, galaxy).

Depending on the number of stimuli required per cluster for individual studies, researchers may also wish to know how many stimuli can safely be sampled from the clusters, in their order of membership certainty. One solution could be to use the criteria from the default Mclust() (Fraley & Raftery, 2006) graphical output in R, which considers images with uncertainties below the 75th percentile to be appropriately clustered. Of course, more conservative cutoffs could be selected, should the amount of data support it, the number of stimuli required be relatively small, or the study imply high stakes (e.g., in clinical research).

If, on the other hand, researchers require larger numbers of images per cluster than, for instance, those having uncertainties below the 75th percentile, or even more than the size of the smallest clusters extracted (e.g., $N = 71$, in our case), several solutions exist. First, one can relax the reliance on uncertainties when excluding images, but nevertheless retain the uncertainties for use as statistical weights in models, after experimental data have been collected. This would ensure that better cluster representatives would count more when determining the research results, making images with higher uncertainties still usable. A second alternative could be to resort to sampling additional photographic stimuli from other databases. To the extent that PAD ratings/norms exist or can be obtained for such images, it would be trivial to determine their cluster memberships with regard to the present results.

Finally, it is also possible for researchers to modify our method to suit their aims—for instance, in terms of the criteria used for the CI widths, or the level of uncertainty used to determine clear cluster memberships—as long as there is good justification for doing so and deviating from the standard approach (e.g., in clinical research with high stakes).

A comparison between our method and ad-hoc approaches to selecting IAPS stimuli

On the basis of our brief comparison, we discovered that a common practice is to group together stimuli that, according to our method, actually represent different types of negative or positive images (e.g., when a single group of positive material is used, instead of one positive cluster of "serene scenes," with lower arousal and somewhat higher dominance, plus one cluster of "exciting scenes," with higher arousal and somewhat lower dominance). Thus, a single, generic grouping of "positive" (or "negative") images may obscure any specific effects due to just one *type of* positive (or negative) material—particularly if the effects actually differ between the several types of positive (or negative) images.

This would be in addition to the relatively frequent inclusion of outliers in the literature, and importantly, of less reliable images (with 95 % CIs wider than one point). Of these, outliers could be ethically risky, and should be avoided especially when relying on cluster analysis for stimulus selection (otherwise, they may distort the clustering solutions), whereas images with wide CIs can introduce additional error variance into research results.

Another interesting finding that emerged from our comparison is that effects can become diluted if neutral categories are not truly neutral, and extend into the space of clusters that we have found to actually be mildly positive or negative. This could result in diminished power to detect differences between the "neutral" and positive or negative stimulus categories.

Finally, we would underline that we do not wish to highlight these differences as criticisms of previous research using the IAPS. Rather, it is our intention to improve on these very widespread methods for selecting stimuli, by promoting our novel method that relies on model-based cluster analysis. Indeed, we believe previous image selection techniques may still be useful in limited contexts; however, it would be very difficult to predict when or to what extent they might influence results (by obscuring effects or "diluting" them, etc.). In addition, they may often vary considerably from study to study (in terms of both selection criteria and resulting selections), making comparisons between studies more difficult. As such, we argue that relying on a statistical, easily reproducible[9] and automatic procedure, which also quantifies the extent to which images belong to a given cluster, is much to be preferred.

Further research and limitations

Despite being arguably more objective than "manual" selection methods, cluster analysis is not an "exact science." As has been shown previously, the large variety of algorithms available can lead to substantial variations in clustering solutions. It is sometimes partly up to the researcher to decide which clustering solution is appropriate for their data. This is particularly the case with *k*-means and hierarchical clustering, because the clustering process is initialized using random seeds and/or various clustering indices that may suggest conflicting numbers of clusters. In contrast, with model-based clustering such difficulties can largely be avoided, because the results are identical on different runs of the algorithm (unlike *k*-means), and the only relevant criterion for choosing the number of clusters is the BIC.

Thus, any flexibility attributed to clustering methods (model-based clustering, in particular) may be seen as an asset, rather than a risk for objectivity, as long as the choices made by researchers (i.e., level of uncertainty, the width of CIs, etc.) are transparent and justified by convincing arguments. The present work aims only to provide a guide for a method that is more appropriate than manual selection strategies—particularly if multiple dimensions are used simultaneously for selecting stimuli.

In addition, although the cases sampled from each cluster acquit themselves of being good cluster representatives, the overall selection of treatment levels (or clusters) is ultimately constrained by the type of data in the IAPS—or whichever stimulus database would be used in research. As such, the final selection of stimuli cannot include categories of stimuli that are not part of the database to begin with. In the case of IAPS data, this may be either because such stimuli would be difficult to find, due to the PAD correlations (e.g., very

[9] As long as any researchers using model-based clustering are transparent about all of the settings/data-cleaning methods used with the procedure.

negative images with low arousal are unlikely), or because the IAPS domain of images does not include emotional material that extends as far as possible within the 3-D PAD space (e.g., images with moderate valence and moderate, rather than low, arousal are not very common).

These concerns could be addressed in the future either by the inclusion of new images or by a renorming process for the IAPS database (potentially via Amazon Mechanical Turk), using larger samples to rate each image. This can also present the added benefit of the average values being more stable (i.e., smaller standard deviations), and therefore fewer images being filtered out of the clustering procedure, thus creating more comprehensive clusters. However, until then, when interpreting results based on the current IAPS norms, the empty areas in the PAD space will require careful consideration, since otherwise research conclusions may be biased.

In terms of future research, an interesting avenue would be to compare empirical results when using a manual image selection method, relative to our cluster-analysis-based classification. Also, there is room yet for further standardization of the IAPS images—for example, in terms of their spatial frequency content (i.e., their level of detail or "coarseness"), which may interact with their affective processing (Delplanque, N'diaye, Scherer, & Grandjean, 2007). Cluster analysis could take such dimensions (as well as participant age, etc.) into account when creating experimental treatment levels, provided they have been converted to standard scores beforehand. Furthermore, depending on whether the raw data used to produce the IAPS normative ratings will be made available, the source of the large standard deviations could be explored further, to indicate improved selection strategies.

Finally, for any research requiring "emotionally ambiguous" stimuli, which do not clearly fit into any particular cluster, uncertainty estimates for the classification of images may provide a more empirically principled means to identify these along multiple dimensions. This would represent a higher level of rigor, the application of which could be explored in future research.

Conclusions

In this article, we have presented a method for selecting experimental stimuli, which we have illustrated using the IAPS database. Using model-based clustering and valence arousal, and dominance scores, we classified the IAPS images into five categories—with each image presenting a certain level of certainty of belonging to its respective cluster. Our method is flexible, efficient, and reproducible, and it provides meaningful clusters in a symmetrical format, in terms of their valence ratings: two negative clusters (one more so than the other); one neutral cluster; and two positive clusters (one more so than the other). However, this method could easily be extended to other stimulus databases, in which the same principles may be applied: careful data inspection, including the removal of any duplicated cases in the stimulus database; the exclusion of missing values and outliers (in a judicious manner); selecting the most precise cases; selecting an appropriate clustering algorithm and clustering solution; and finally, extracting a constant number of stimulus exemplars from each cluster.

Acknowledgments The first author thanks Marian Popa, former senior statistics lecturer at the University of Bucharest, for all of his advice over the past few years, as well as some suggestions for this article. All of the authors thank Sara Dolnicar, who kindly provided input on a measure referenced in our work.

References

Aguilar de Arcos, F., Verdejo-García, A., Peralta-Ramírez, M. I., Sánchez-Barrera, M., & Pérez-García, M. (2005). Experience of emotions in substance abusers exposed to images containing neutral, positive, and negative affective stimuli. *Drug and Alcohol Dependence, 78,* 159–167.

Anderson, E., Siegel, E. H., & Barrett, L. F. (2011). What you feel influences what you see: The role of affective feelings in resolving binocular rivalry. *Journal of Experimental Social Psychology, 47,* 856–860.

Ball, G. H., & Hall, D. J. (1965). *ISODATA, a novel method of data analysis and pattern classification (Technical report).* Menlo Park, CA: Stanford Research Institute.

Bernat, E., Patrick, C. J., Benning, S. D., & Tellegen, A. (2006). Effects of picture content and intensity on affective physiological response. *Psychophysiology, 43,* 93–103.

Borcard, D., Gillet, F., & Legendre, P. (2011). *Numerical ecology with R*. New York, NY: Springer.

Bradley, M. M., & Lang, P. J. (1994). Measuring emotion: The self-assessment manikin and the semantic differential. *Journal of Behavior Therapy and Experimental Psychiatry, 25,* 49–59.

Bradley, M. M., & Lang, P. J. (2007). The International Affective Picture System (IAPS) in the study of emotion and attention. In J. A. Coan & J. J. Allen (Eds.), *Handbook of emotion elicitation and assessment, Series in affective science* (pp. 29–46). New York, NY: Oxford University Press.

Caliński, T., & Harabasz, J. (1974). A dendrite method for cluster analysis. *Communications in Statistics—Theory and Methods, 3,* 1–27.

Deen, B., Pitskel, N. B., & Pelphrey, K. A. (2011). Three systems of insular functional connectivity identified with cluster analysis. *Cerebral Cortex, 21,* 1498–1506. doi:10.1093/cercor/bhq186

Delplanque, S., N'diaye, K., Scherer, K., & Grandjean, D. (2007). Spatial frequencies or emotional effects? A systematic measure of spatial frequencies for IAPS pictures by a discrete wavelet analysis. *Journal of Neuroscience Methods, 165,* 144–150.

Dimitriadou, E., Dolničar, S., & Weingessel, A. (2002). An examination of indexes for determining the number of clusters in binary data sets. *Psychometrika, 67,* 137–159.

Do, C. B., & Batzoglou, S. (2008). What is the expectation maximization algorithm*? Nature Biotechnology, 26,* 897–899.

Dolnicar, S., Grabler, K., & Mazanec, J. A. (1999). A tale of three cities: Perceptual charting for analyzing destination images. In A. G. Woodside, G. I. Crouch, J. A. Mazanec, M. Oppermann, & M. Y.

Sakai (Eds.), *Consumer psychology of tourism, hospitality and leisure* (pp. 39–62). Wallingford, UK: CAB International.

Eizenman, M., Yu, L. H., Grupp, L., Eizenman, E., Ellenbogen, M., Gemar, M., & Levitan, R. D. (2003). A naturalistic visual scanning approach to assess selective attention in major depressive disorder. *Psychiatry Research, 118,* 117–128.

Fraley, C., & Raftery, A. E. (2002). Model-based clustering, discriminant analysis, and density estimation. *Journal of the American Statistical Association, 97,* 611–631.

Fraley, C., & Raftery, A. E. (2006). *MCLUST version 3: An R package for normal mixture modeling and model-based clustering* (Report No. 504). Seattle, WA: University of Washington, Dept. of Statistics.

Glenn, C. R., Blumenthal, T. D., Klonsky, E. D., & Hajcak, G. (2011). Emotional reactivity in nonsuicidal self-injury: Divergence between self-report and startle measures. *International Journal of Psychophysiology, 80,* 166–170.

Grühn, D., & Scheibe, S. (2008). Age-related differences in valence and arousal ratings of pictures from the International Affective Picture System (IAPS): Do ratings become more extreme with age? *Behavior Research Methods, 40,* 512–521.

Hallahan, M., & Rosenthal, R. (1996). Statistical power: Concepts, procedures, and applications. *Behaviour Research and Therapy, 34,* 489–499.

Hamann, S. B., Ely, T. D., Hoffman, J. M., & Kilts, C. D. (2002). Ecstasy and agony: Activation of the human amygdala in positive and negative emotion. *Psychological Science, 13,* 135–141.

Hartigan, J. A. (1975). *Clustering algorithms*. New York, NY: Wiley.

Hartigan, J. A., & Wong, M. A. (1979). Algorithm AS 136: A *k*-means clustering algorithm. *Journal of Applied Statistics, 28,* 100–108.

Hautamäki, V., Cherednichenko, S., Kärkkäinen, I., Kinnunen, T., & Fränti, P. (2005). Improving *k*-means by outlier removal. In *Proceedings of the 14th Scandinavian Conference on Image Analysis* (pp. 978–987). Berlin, Germany: Springer-Verlag.

Hlavac, M. (2013). *stargazer: LaTeX code and ASCII text for well-formatted regression and summary statistics tables* (version 4.5.3) [R package and documentation]. Retrieved on June 20, 2015, from CRAN.R-project.org/package = stargazer.

Hornik, K. (2005). A CLUE for CLUster ensembles. *Journal of Statistical Software, 14,* 1–25.

Hubert, L., & Arabie, P. (1985). Comparing partitions. *Journal of Classification, 2,* 193–218.

Ito, T. A., Cacioppo, J. T., & Lang, P. J. (1998). Eliciting affect using the International Affective Picture System: Trajectories through evaluative space. *Personality and Social Psychology Bulletin, 24,* 855–879.

Jain, A. K. (2010). Data clustering: 50 years beyond *K*-means. *Pattern Recognition Letters, 31,* 651–666.

Jain, A. K., & Dubes, R. C. (1988). *Algorithms for clustering data* (Vol. 6). Englewood Cliffs: Prentice Hall.

Kantardzic, M. (2011). *Data mining: Concepts, models, methods, and algorithms* (2nd ed.). Hoboken, NJ: Wiley.

Kaufman, L., & Rousseeuw, P. J. (2005). *Finding groups in data: An introduction to cluster analysis*. Hoboken, NJ: Wiley.

Koenigsberg, H. W., Fan, J., Ochsner, K. N., Liu, X., Guise, K., Pizzarello, S., & New, A. (2010). Neural correlates of using distancing to regulate emotional responses to social situations. *Neuropsychologia, 48,* 1813–1822.

Lane, R. D., Reiman, E. M., Bradley, M. M., Lang, P. J., Ahern, G. L., Davidson, R. J., & Schwartz, G. E. (1997). Neuroanatomical correlates of pleasant and unpleasant emotion. *Neuropsychologia, 35,* 1437–1444.

Lang, P. J. (1980). Technology in mental health care delivery systems. In J. B. Sidowski, J. H. Johnson, & T. A. Williams (Eds.), *Behavioral treatment and bio-behavioral assessment: Computer applications* (pp. 119–l37). Norwood, NJ: Ablex.

Lang, P. J., Bradley, M. M., & Cuthbert, B. N. (2008). *International affective picture system (IAPS): Affective ratings of pictures and instruction manual* (Report No. A-8). Gainesville, FL: University of Florida, NIMH Center for the Study of Emotion and Attention.

LeDoux, J. E. (1996). *The emotional brain*. New York: Simon & Shuster.

Leys, C., Ley, C., Klein, O., Bernard, P., & Licata, L. (2013). Detecting outliers: Do not use standard deviation around the mean, use absolute deviation around the median. *Journal of Experimental Social Psychology, 49,* 764–766.

Lithari, C., Frantzidis, C., Papadelis, C., Vivas, A. B., Klados, M., Kourtidou-Papadeli, C., & Bamidis, P. (2010). Are females more responsive to emotional stimuli? A neurophysiological study across arousal and valence dimensions. *Brain Topography, 23,* 27–40.

MacQueen, J. (1967). Some methods for classification and analysis of multivariate observations. In L. Lecam & J. Neyman (Eds.), *Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability* (Vol. 1, pp. 281–297). Berkeley, CA: The Regents of the University of California.

Mehrabian, A. (1996). Pleasure-arousal-dominance: A general framework for describing and measuring individual differences in temperament. *Current Psychology, 14,* 261–292.

Meilă, M. (2007). Comparing clusterings—An information based distance. *Journal of Multivariate Analysis, 98,* 873–895.

Mikels, J. A., Fredrickson, B. L., Larkin, G. R., Lindberg, C. M., Maglio, S. J., & Reuter-Lorenz, P. A. (2005). Emotional category data on images from the International Affective Picture System. *Behavior Research Methods, 37,* 626–630.

Mikels, J. A., Larkin, G. R., Reuter-Lorenz, P. A., & Carstensen, L. L. (2005). Divergent trajectories in the aging mind: Changes in working memory for affective versus visual information with age. *Psychology and Aging, 20,* 542–553.

Most, S. B., Chun, M. M., Widders, D. M., & Zald, D. H. (2005). Attentional rubbernecking: Cognitive control and personality in emotion-induced blindness. *Psychonomic Bulletin & Review, 12,* 654–661.

Mun, E. Y., von Eye, A., Bates, M. E., & Vaschillo, E. G. (2008). Finding groups using model-based cluster analysis: Heterogeneous emotional self-regulatory processes and heavy alcohol use risk. *Developmental Psychology, 44,* 481–495. doi:10.1037/0012-1649.44.2.481

Perri, R. L., Berchicci, M., Lucci, G., Cimmino, R. L., Bello, A., & Di Russo, F. (2014). Getting ready for an emotion: Specific premotor brain activities for self-administered emotional pictures. *Frontiers in Behavioral Neuroscience, 8,* 197.

R Development Core Team. (2015). R: A language and environment for statistical computing (Version 3.2.1) [Computer software]. Vienna, Austria: R Foundation for Statistical Computing. Retrieved from: www.R-project.org

Russell, J. A., & Mehrabian, A. (1977). Evidence for a three-factor theory of emotions. *Journal of Research in Personality, 11,* 273–294.

Stins, J. F., & Beek, P. J. (2007). Effects of affective picture viewing on postural control. *BMC Neuroscience, 8,* 83.

Tomaszczyk, J. C., Fernandes, M. A., & MacLeod, C. M. (2008). Personal relevance modulates the positivity bias in recall of emotional pictures in older adults. *Psychonomic Bulletin & Review, 15,* 191–196.

Watson, D., Clark, L. A., & Tellegen, A. (1988). Development and validation of brief measures of positive and negative affect: The PANAS scales. *Journal of Personality and Social Psychology, 54,* 1063–1070.

Wu, J. (2012). *Advances in k-means clustering: A data mining thinking*. Berlin: Springer Verlag.

Xing, C., & Isaacowitz, D. M. (2006). Aiming at happiness: How motivation affects attention to and memory for emotional images. *Motivation and Emotion, 30,* 243–250.

Xu, R., & Wunsch, D. (2009). *Clustering* (Vol. 10). New Jersey: John Wiley & Sons.

Zhang, X., Hui, W. Y., & Barrett, L. F. (2014). How does this make you feel? A comparison of four affect induction procedures. *Frontiers in Psychology, 5,* 689.

Zhao, X., & Sandelin, A. (2012). GMD: Measuring the distance between histograms with applications on high-throughput sequencing reads. *Bioinformatics, 28,* 1164–1165.

H.2 Behavior Research Methods supplementary material

A Cluster-Based Approach to Selecting Representative Stimuli from the International Affective Picture System (IAPS) Database

- Supplementary material -

Alexandra C. Constantinescu, Maria Wolters, Adam Moore and Sarah E. MacPherson

Preliminary analyses

Representativeness / precision of measures. A further screen – the coefficient of variation (*cv*)[1] – was used alongside the 95% confidence intervals to assess how representative the normative ratings are for the sample distributions they were computed from. Normally, a measure of variability tailored to ordinal data would have been most appropriate here, such as the Index of Ordinal Variation, or the Coefficient of Ordinal Variation (see Berry & Mielke, 1992; Blair & Lucy, 2000; Kvalseth, 1995). However, its application is not possible without the raw IAPS data forming the norms - which are not made available. Thus, the *cv* was used instead, as it relies only on the IAPS norms/means and standard deviations provided.

We aimed to exclude stimuli with unrepresentative means (i.e., with $cv > 30$, Brown, 1998), but this measure proved to be far too conservative for the IAPS dataset, given that only one

---

[1] $cv = (sd / m) \times 100$. The *cv* usually requires a *ratio* scale with an absolute 0 value, to allow for scale transformations. While the nine-point PAD scales can be seen as *interval* scales with adequate measurement properties (Preston & Colman, 2000), they lack an absolute 0. Regardless, because in practice the IAPS scales are not converted to any other system of measurement, using the *cv* on IAPS data was considered an acceptable step.

case complied with this criterion simultaneously for all three PAD scales, out of the N=849. Consequently, only the criteria based on outlying values and confidence intervals were maintained, as otherwise using the coefficient of variation would have resulted in the exclusion of nearly all stimuli, except one.

Results

Clustering techniques

K-means clustering. Euclidean distances were chosen to create the k partitions[2], and various clustering indices were calculated to assess the suitability of extracting between two and eight clusters. These indices were offered by the clustIndex() function from the cclust R package (Dimitriadou, 2014), and included: the *Calinski-Harabasz Index* (Caliński & Harabasz, 1974), the *Ball Index* (Ball & Hall, 1965), the *Hartigan Index* (Hartigan, 1975) and the *Simple Structure Index* (SSI, Dimitriadou et al., 2002; Dolnicar et al., 1999). The first three of these are based on within-/between-cluster sums of squares calculations (i.e., minimizing the former and/or maximizing the latter to ensure cluster compactness and/or separation between clusters), whereas the fourth is a composite measure taking into account: the maximum difference between clusters, the size of the most different clusters, and the difference between cluster centroids and the grand mean of each dimension. For details, please refer to Table S1.

[2] Although other options exist for computing distances between data points, Euclidean distances are the most common for *k*-means, particularly as this clustering method was developed with them in mind (Hartigan & Wong, 1979; Jain, 2010).

In addition, we consulted the clValid R package (Brock et al., 2008) in order to assess the optimal value of *k* in terms of internal validation measures. Thus, using Euclidean distances, we computed: the *Dunn Index* (Dunn, 1973), i.e., the ratio between the smallest distance between cases assigned to different clusters, and the maximum distance between two cases assigned to the same cluster; the *measure of Connectivity* (Handl et al., 2005), i.e., the extent to which neighboring data points share cluster membership; and the *Average Silhouette Width* (ASW, Rousseeuw, 1987), i.e., the extent to which data points are closer to other points in their own cluster, rather than the nearest points assigned to a different cluster. For details on these indices, please refer to Table S1. Overall, the Connectivity, Average Silhouette Width, Calinski-Harabasz Index and Simple Structure Index pointed to a lower number of clusters: two (in the case of the former two indices) or three (the latter two indices), whereas the Hartigan, Ball and the Dunn Indices, all pointed to eight clusters, the maximum number tested.

Finally, we also used the NbClust R package (Charrad et al., 2014) which computes thirty such indices and indicates which value of *k* is considered to be optimal, of those tested. Thus, when testing each option between k=2 and k=15, twenty of the clustering indices recommended to extract either two or three clusters, with the remaining indices suggesting that four, ten, fourteen or even fifteen clusters (the maximum tested) be extracted. In order to observe the behaviour of the algorithms, we also extended the number of clusters tested to thirty, rather than fifteen, and re-assessed the clustering indices. In this case, the majority of indices (i.e, seventeen out of thirty) again suggested to extract either two or three clusters from the data, with the remaining indices recommending four, ten, fifteen, seventeen or even thirty (again, the maximum number tested).

Thus, it is apparent that two diverging trends exist within the IAPS data: on the one hand, a tendency to group a large number of images in just two or three clusters, and on the other, a tendency to achieve a finer-grained grouping - although in this case, the number of clusters can vary considerably by the index used, with one or two clustering indices endorsing each of: four, ten, fourteen, fifteen, seventeen or even thirty clusters in the data, depending on the maximum number allowed in the search.

Hierarchical clustering. This is also a hard clustering method, where each case is assigned to exclusively one cluster, rather than being assigned a probability of membership. Various combinations of linkage methods and distance metrics were used to identify one clustering solution which most correlated with the original distance matrix (i.e., a matrix containing distances between every pair of cases). The possible linkage methods include: Single, Complete, Average Linkage, and the Ward method. Each of these strategies differs in the decision criterion for progressively merging clusters, i.e., based on the distance between the closest points belonging to different clusters (*Single Linkage*), on the furthest apart such points (*Complete Linkage*), on the average distance between points in two clusters (*Average Linkage*), or on which merge would minimize within-cluster variance (*Ward method*, which is similar to *k*-means in that both are least-squares methods; Borcard et al., 2011).

Testing various linkage methods and distance metrics jointly allowed us to find the combination yielding the clustering solution with the highest degree of similarity (i.e., the cophenetic correlation, Jain et al., 1988; Rohlf & Fisher, 1968; Xu & Wunsch, 2009) to the original data. In other words, the solution that has minimal dissimilarity to the raw data.

Due to the relationships between valence, arousal, and dominance displayed in Figure 1 in the main article, distances based on correlations were also used, in addition to Euclidean distances. Each of these was combined with one of the four agglomeration methods, and out of the resulting eight possible combinations, the weakest cophenetic correlation was observed for **Single Linkage** combined with Euclidean distances, r = 0.22, whereas in combination with correlation distances, Single Linkage produced a cophenetic correlation of r = 0.87. **Complete Linkage**, on the other hand, produced slightly lower correlations with r = 0.72 and r = 0.77 for Euclidean and correlation distances, respectively. **Ward Linkage** with Euclidean distances (which bears resemblance to k-means) produced a cophenetic correlation of r = 0.72, and r = 0.82 with correlation distances. Finally, **Average Linkage** paired with correlation-based distances produced the most similar results to the original distance matrix, reaching the maximum cophenetic correlation of r = 0.91, whereas with Euclidean distances, the correlation dropped to r = 0.71. Thus, Average Linkage paired with correlation distances represents the most suitable combination for the IAPS data. The Gower distance (a measure of dissimilarity, Gower, 1971) was also minimized for this same combination: 40030.17, relative to the next best value, achieved by Complete Linkage with correlation distances (190072.7). Single and Ward Linkage both performed more poorly on to this measure.

After having identified the most suitable hierarchical agglomeration method for this dataset, we proceeded to use the clValid R package (Brock et al., 2008) to again determine the most appropriate number of clusters in the data. This package conveniently offers the same combination of agglomeration method and distance metric which was proven to be most suitable for IAPS data.

In terms of both Connectivity and Average Silhouette Widths (briefly described above), a number of two clusters was suggested, whereas the Dunn Index indicated three. Mantel optimality (the correlation between the original distance matrix and binary matrices showing cluster membership; Borcard et al., 2011) also suggested two. However, as with *k*-means, there is some variability to be found, as when using the Elbow method for partitioning variance into clusters (GMD R package; Zhao & Sandelin, 2012), the optimal number of clusters (also based on Average Linkage) indicated was seven.

Finally, using the NbClust package (Charrad et al., 2014) with Average Linkage and correlation distances, and testing a maximum number of fifteen clusters, the most endorsed options (by sixteen indices overall) were two or nine clusters, with fewer other indices suggesting each of: three, four, twelve, or fifteen to extract. When extending the search limit to thirty clusters to extract, two and nine clusters remained the most frequent recommendation (from sixteen indices), with a few other indices suggesting three, four, five, fifteen, twenty-nine, or thirty clusters (the maximum number tested). Thus, similarly to *k*-means, the same tendency is apparent for hierarchical clustering: either a very small, heterogeneous number of clusters is endorsed, or larger numbers of small, fine-grained clusters emerges instead. The notable difference is that, unlike for *k*-means, the nine-cluster (rather than the three-cluster) solution is endorsed much more heavily for hierarchical clustering, as computed with the NbClust package.

Model-based clustering. We verified whether the assumptions for model-based clustering were satisfied (i.e., multivariate normality for each component distribution/cluster). To this end, we computed a mixture model using model-based clustering, as implemented within the R package mclust (Fraley & Raftery, 2006). We subsequently examined the eigenvalues and the densities of principal component scores associated with each component in the mixture model. Eigenvalues describe the amount of variation that characterizes each axis in the 3D space where the clusters are defined: ellipsoidal shapes are consistent with normal multivariate distributions, and exhibit larger variations along one axis, and smaller for the other two – as shown in Table S2. Principal component scores largely formed bell-curve shapes for each cluster, suggesting a normal/symmetrical distribution of cases within the 3D space.

Further support for the normality assumption is provided by other multivariate normality tests also listed in Table S2, as well as the 3D density plots for each cluster (or each mixture component), in Figure S1. Thus, we conclude that the conditions of use for model-based clustering were generally satisfied. A variety of model-based clustering models were therefore computed for the data, and their associated BIC values are provided in Table S3. Finally, the BIC-optimal classification of cases (presented alongside uncertainties) is presented in the repository at: www.github.com/CaterinaC/IAPSClustering2016.

Validating the clustering solutions

Finding a stable structure within the data, across methods. Assuming that the IAPS data presents a clear, discernible structure, all clustering algorithms should in principle be able to identify this structure despite computational differences. In order to check this, we assessed the extent to which model-based clustering yields membership assignments that overlap with those from the other two competing methods.

When using the Adjusted Rand Index (ARI, a conservative measure which penalizes for any randomness in the overlap, Hubert & Arabie, 1985), the degrees of association between the model-based solution and the other two solutions were 0.474 and 0.318, for *k*-means and hierarchical clustering, respectively. The association with an entirely random classification of IAPS cases predictably dropped to ≈ 0. As Steinley (2004) considers ARI values greater than 0.90 - excellent, values greater than 0.80 - good, values greater than 0.65 - moderate, and values less than 0.65 - poor, the values observed for ARI in our dataset seem to indicate each method creates partitions with relatively little in common with the other two.

We next computed Meilă's (2007) variation of information (VI) criterion (implemented in R package igraph, function compare.communities(); Csardi & Nepusz, 2006), which measures (in bits) how much information is lost or gained by moving from one partitioning to another. This amount should be small if $k = 5$ is an adequate structure for the IAPS data, and is picked up similarly by the various clustering algorithms. The overlap (in terms of VI) between model-based clustering and *k*-means classification is 1.187, and between model-based clustering and hierarchical clustering is 1.462. In order for the overlap to have a possible range between 0 and 1 (where 0 would indicate identical clustering solutions across methods), we normalized values by dividing them with the log of the sample size used (Meilă, 2007). This converted the indices to 0.176 and 0.217, respectively, which can be considered somewhat low values, suggesting there is not much information to be gained/lost when moving from one classification to another, and that there is enough similarity between partitions of five clusters, regardless of the algorithm used to produce them. For details on these measures, please refer to Table S1.

Finally, we used Cramer's ϕ to assess pairwise overlap between the five-cluster classifications from all three methods. Cross-tabulating *k*-means and model-based classifications led to

Cramer's ϕ = 0.704 (a strong association), whereas crossing model-based and hierarchical classifications led to a lower Cramer's ϕ = 0.516 (but still indicating a relatively strong association; Kotrlik et al., 2011). Though the ARI results lend some ambiguity, on the whole these results constitute moderate evidence that a specific, five-cluster data structure can be identified in the IAPS, given the level of agreement between the clustering methods.

Comparing the fit of the model-based clustering solution to that of the other two algorithms. As a further test, predictive models were built using the interaction term between valence, arousal, and dominance scores, or each of these dimensions on its own, as the outcome, and one of the five-component clustering solutions (*k*-means, hierarchical, and model-based) as the predictor, in order to compare R^2 between them, and thus gauge which clustering method yields groups that most accurately reflect back the original data. This approach resulted in a total of twelve predictive models (four outcomes by three clustering methods). In all cases, models were significant, with R^2 ranging from 0.430 to 0.885, and an average across all methods of R^2 = 0.724 (*SD* = 0.130). This indicates that, on average, 72.4% of the variation seen in outcomes (i.e. raw data) was explained by membership assignment to the five clusters, which is considerable. Strictly referring to model-based clustering, this method achieved R^2 = 0.718 for predicting the interaction term, and R^2 = 0.885 for predicting Valence scores only, R^2 = 0.430 for Arousal and R^2 = 0.744 for Dominance.

In addition, using an additive model building strategy, nested models were compared using the anova() function in R, to verify if each clustering method could explain more variation in the outcomes, above and beyond previously inserted predictors (i.e., clustering solutions). We found that model-based clustering is a complementary method to the other two methods, as each method helped explain significantly more variance in outcomes when added to a model containing one other clustering method. For instance, a model predicting the Valence × Arousal × Dominance interaction based on *k*-means or hierarchical clusters would be improved significantly by the addition of a predictor coding model-based clusters, but also vice versa. Thus we were able to roughly recreate the original data using complementary solutions (i.e., variance explained is not completely overlapping): across all pairwise combinations of predictors, the average R^2 achieved was 0.814 (*SD*=0.016). Specifically, when adding *k*-means or hierarchical clustering classifications to a model including only the model-based classification as a predictor, the boost in R^2 was of 0.11 and 0.08, respectively (significant in both cases at $p < 0.001$). Conversely, when adding the model-based classification as a predictor alongside either of the other two classifications, the rise in R^2 is of 0.02 and 0.20, respectively ($p < 0.001$, in either case).

Evaluating the stability of the model-based clustering solution. The IAPS data were divided into two randomly selected halves in order to assess split-half validation. The two halves were arbitrarily defined as *the learning dataset* (used to create a model-based clustering solution with five components), and *the test dataset* (whose clustering solution could be predicted using the solution generated for the learning dataset). If the five-cluster structure found using model-based clustering is appropriate for the IAPS data, then the prediction for the test dataset, and a clustering solution created independently for the test dataset, should match closely.

This procedure was repeated twice, with each random half of the dataset being considered as the learning dataset on a given run. As such, both random halves of the data were subject to model-based clustering, with the specification of extracting five clusters (which was the optimal solution for the dataset considered in its entirety). Subsequently, using the cl_predict()

R function from package clue (Hornik, 2005), each classification was used to predict the clustering structure of the other half (the test data).

The degree of association between the predicted clustering of the test dataset and its actual clustering was quantified using a measure of effect size for their cross-tabulation. The average value computed for ϕ Cramer = 0.864, which is a very strong association (Kotrlik et al., 2011). Given this result, the five-cluster solution provided by Mclust() was judged to be robust and well supported by the data.

In addition, in order to reassess how stable the k = 5 solution issued by Mclust() was, and/or how much of it was potentially due to multivariate outlying cases, we used a jack-knife procedure to randomly remove 10% of the cases from the dataset, during 7500 bootstrap repetitions[3]. Ideally, if the data structure is represented consistently by the clusters, then no major changes should occur with reference to the optimal number of k. After each repetition, the optimal value for k was reassessed, and across all repetitions, aggregated data suggest the most commonly occurring optimal solution for model-based clustering when 10% of cases were removed was k = 3 (46.67%), followed by k = 4 (34.67%), and finally, k = 5 (18.67%).

Given the departure from the k = 5 Mclust() solution achieved on the full dataset, we assessed if there were notable discrepancies between the k = 3 and k = 5 solutions achieved on the random subsets of the data. Therefore, using 2500 repetitions, on each random data subset we fitted both a three-component and a five-component mixture model using Mclust(), and then

[3] Due to limitations in computing power, this number will vary across the various procedures used to validate the clustering solution.

assessed the (mis-)match between them. Collectively, the 2500 resulting cross-tabulations achieved an average Cramer's $\phi = 0.92$, with a min = 0.81 and max = 0.99. Therefore the differences seen in the value of *k* most likely reflect the fact that one or two clusters from the *k*=5 solution were collapsed due to the induced data attrition (-10%), but that similarities between solutions nevertheless remained robust.

In order to investigate any further the differences between methods, a similar approach was adopted for *k*-means and hierarchical clustering, by randomly removing 10% of the data during 1000 repetitions. Euclidean distances were used for *k*-means, with correlation distances and Average Linkage being specified for hierarchical clustering. For each of these clustering methods and during each repetition, the adequate number for *k* was recomputed based on measures of Connectivity, Average Silhouette Width, and Dunn Index, offered by the clValid() R function (within package clValid by Brock et al., 2008). After 10% of the values had been removed, 82.80% and 71.63% of the time, only one cluster emerged for *k*-means and hierarchical clustering, respectively (i.e., no cluster structure could be defined). Given that the variability in the dataset leads to ≈ 60% of data points having Mahalanobis distances larger than 2 (and peaking at 19.559), extracting one single cluster was judged to be an unrealistic solution, and symptomatic of an inefficient clustering process for this particular dataset. This suggests that model-based clustering is more appropriate for IAPS data, relative to *k*-means or hierarchical clustering.

References

Ball, G. H. & Hall, D. J. (1965). *ISODATA, a novel method of data analysis and pattern classification* (Technical report). Menlo Park, CA: Stanford Research Institute.

Berry, K. J., & Mielke Jr, P. W. (1992). Indices of ordinal variation. *Perceptual and Motor Skills*, *74*(2), 576-578.

Blair, J., & Lacy, M. G. (2000). Statistics of ordinal variation. *Sociological Methods and Research*, *28*(3), 251-280.

Borcard, D., Gillet, F., & Legendre, P. (2011). *Numerical ecology with R*. New York: Springer

Brock, G., Pihur, V., Datta, S., & Datta, S. (2008). clValid: An R package for cluster validation. *Journal of Statistical Software, 25*(4), 1-22. Retrieved from: http://www.jstatsoft.org/v25/i04/paper.pdf, on August 14, 2013.

Brown, C. (1998). *Applied multivariate statistics in geohydrology and related sciences.* Berlin: Springer Science & Business Media.

Caliński, T. & Harabasz, J. (1974). A dendrite method for cluster analysis. *Communications in Statistics - Theory and Methods*, *3*(1), 1–27.

Charrad, M., Ghazzali, N., Boiteau, V., & Niknafs, A. (2014). NbClust: an R package for determining the relevant number of clusters in a data set. *Journal of Statistical Software*, *61*(6), 1-36.

Csardi, G. & Nepusz, T. (2006). The igraph software package for complex network research. *InterJournal Complex Systems, 1695*(5), 1–9.

Dimitriadou, E., Dolničar, S., & Weingessel, A. (2002). An examination of indexes for determining the number of clusters in binary data sets. *Psychometrika*, *67*(1), 137–159.

Dimitriadou, E. (2014). *cclust: Convex Clustering Methods & Clustering Indexes* (version 0.6-19) [R package and documentation]. Retrieved from: http://CRAN.R-project.org/package=cclust, on June 20, 2015

Dolnicar, S., Grabler, K., Mazanec, J. A. (1999). A tale of three cities: Perceptual charting for analysing destination images. In A. G. Woodside, G. I. Crouch, J. A. Mazanec, M. Opperman, & M. Y. Sakai (Eds.), *Consumer psychology of tourism, hospitality and leisure* (pp. 39-62). Wallingford, UK: CAB International Publishing.

Dunn, J. C. (1973). A fuzzy relative of the ISODATA process and its use in detecting compact well-separated clusters. *Journal of Cybernetics, 3*(3), 32–57.

Fraley, C., & Raftery, A. E. (2006). *mclust version 3: An R package for normal mixture modeling and model-based clustering* (Report No. 504). Seattle, WA: University of Washington, Dept. of Statistics.

C. Fraley, A. E. Raftery, T. B. Murphy and L. Scrucca (2012). *mclust version 4 for R: Normal Mixture Modeling for Model-Based Clustering, Classification, and Density Estimation*. (Report No. 597). Seattle, WA: University of Washington, Dept. of Statistics.

Gower, J. C. (1971). A general coefficient of similarity and some of its properties. *Biometrics, 27*(4), 857-871.

Handl, J., Knowles, J., & Kell, D. B. (2005). Computational cluster validation in post-genomic data analysis. *Bioinformatics, 21*(15), 3201–3212.

Hartigan, J. A. (1975). *Clustering algorithms*. New York: Wiley.

Hornik, K. (2005). A CLUE for CLUster ensembles. *Journal of Statistical Software, 14*(12), 1-25.

Hubert, L. & Arabie, P. (1985). Comparing partitions. *Journal of Classification, 2*(1), 193–218.

Jain, A. K., & Dubes, R. C. (1988). *Algorithms for clustering data* (Vol. 6). Englewood Cliffs: Prentice Hall.

Jarek, S. (2009). *Mvnormtest: Normality test for multivariate variables (version 0.1-9)* [R package and documentation]. Retrieved from: http://cran.r-project.org/web/packages/mvnormtest/index.html, on January 12, 2016

Korkmaz S., Goksuluk D., Zararsiz G. (2014). MVN: An R Package for Assessing Multivariate Normality. *The R Journal*, *6*(2):151-162.

Kotrlik, J. W., Williams, H. A., & Jabor, M. K. (2011). Reporting and interpreting effect size in quantitative agricultural education research. *Journal of Agricultural Education, 52*(1), 132–142.

Kvalseth, T. O. (1995). Coefficients of variation for nominal and ordinal categorical data. *Perceptual and Motor Skills, 80*(3), 843-847.

Lucas, A. (2014). *amap: Another Multidimensional Analysis Package (version 0.8-14)* [R package and documentation]. Retrieved from: *http://CRAN.R-project.org/package=amap*, on January 12, 2016

Mazanec, J. A., & Strasser, H. (2000). *A nonparametric approach to perceptions-based market segmentation: Foundations* (Vol. 1). New York: Springer.

Meilă, M. (2007). Comparing clusterings - An information based distance. *Journal of Multivariate Analysis, 98*(5), 873–895.

Preston, C. C., & Colman, A. M. (2000). Optimal number of response categories in rating scales: reliability, validity, discriminating power, and respondent preferences. *Acta psychologica, 104*(1), 1-15.

Rohlf, F. J. & Fisher, D. R. (1968). Tests for hierarchical structure in random data sets. *Systematic Biology, 17*(4), 407–412.

Rousseeuw, P. J. (1987). Silhouettes: a graphical aid to the interpretation and validation of cluster analysis. *Journal of Computational and Applied Mathematics, 20*(1), 53–65.

Steinley, D. (2004). Properties of the Hubert-Arabie Adjusted Rand Index. *Psychological Methods*, *9*(3), 386.

Xu, R. & Wunsch, D. (2009). *Clustering* (Vol. 10). New Jersey: John Wiley & Sons.

Zhao, X. & Sandelin, A. (2012). GMD: Measuring the distance between histograms with applications on high-throughput sequencing reads. *Bioinformatics, 28*(8), 1164–1165.

Tables

Table S1. Principal clustering indices consulted.

Index name	Formula	Abbreviations	Eq.
Calinski-Harabasz	$\frac{SSB/(k-1)}{SSW/(N-k)}$	SSB = Sum of squares between; SSW = Sum of squares within; k = number of clusters; N = number of data points.	**(1)**
Ball	$\frac{SSW}{k}$		**(2)**
Hartigan	$\log\left(\frac{SSB}{SSW}\right)$		**(3)**
SSI	$SSI = \sum_{j=1}^{M}\left(w_i max_j - w_i min_j\right)$ $wSSI = \frac{SSI}{SSI_{max}}\frac{N_{max \vee min}}{N}$, where $0 < wSSI < 1$		**(4)**
Dunn	$\frac{\min_{C_k, C_l \in \mathcal{C}, C_k \neq C_l}\left(\min_{i \in C_k, j \in C_l} dist(i,j)\right)}{\max_{C_m \in \mathcal{C}} diam(C_m)}$	$\mathcal{C}$ = a particular clustering partition; $diam(C_m)$ = maximum distance between observations in cluster C_m.	**(5)**
Connectivity	$\sum_{i=1}^{N}\sum_{j=1}^{L} x_{i,nn_{i(j)}}$	L = parameter giving the nearest neighbors to use; $nn_{i(j)}$ = j^{th} nearest neighbor of observation *i;* $x_{i,nn_{i(j)}}$ = 0, if *i* and *j* are in the same cluster, and 1/*j* otherwise.	**(6)**

ASW	$\frac{b_i - a_i}{\max(b_i,\ a_i)}$	a_i = the average distance between *i* and all other observations in the same cluster; b_i = the average distance between *i* and the observations in the nearest neighboring cluster.	**(7)**
ARI	$\frac{\sum_{i,j}\binom{n_{ij}}{2} - \left[\sum_i \binom{a_i}{2} \sum_j \binom{b_j}{2}\right] / \binom{n}{2}}{\frac{1}{2}\left[\sum_i \binom{a_i}{2} + \sum_j \binom{b_j}{2}\right] - \left[\sum_i \binom{a_i}{2} \sum_j \binom{b_j}{2}\right] / \binom{n}{2}}$	Formula equivalent to: AdjustedRandIndex $= \frac{RandIndex - ExpectedIndex}{MaximumIndex - ExpectedIndex}$, where: $\binom{n}{2}$ = number of pairs of observations; $\binom{a_i}{2}$ = the number of distinct pairs that can be constructed within rows; $\binom{b_j}{2}$ = the number of distinct pairs that can be constructed within columns.	**(8)**
VI	$\mathcal{H}(\mathcal{C}) + \mathcal{H}(\mathcal{C}') - 2I(\mathcal{C}, \mathcal{C}')$	$I(\mathcal{C}, \mathcal{C}')$ = mutual information between two clusterings; $\mathcal{H}(\mathcal{C})$ = the entropy associated with clustering $\mathcal{C}$.	**(9)**

Note. For **(1)**, **(2)**, **(3)**, see e.g., Dimitriadou, Dolničar & Weingessel (2002). For **(4)**, see Mazanec & Strasser (2000). For **(5)**, **(7)**, **(6)**, see Brock et al. (2008). For **(8)**, see Hubert & Arabie (1985). For **(9)**, see Meila (2007).

Table S2. Results from various multivariate normality tests run in R, using packages MVN (Korkmaz et al., 2014), mvnormtest (Jarek, 2009) and amap (Lucas, 2014). No cluster is shown to be non-normal by all of the tests simultaneously.

Test	Measure & Value		Cluster 1	Cluster 2	Cluster 3	Cluster 4	Cluster 5
Mardia	Skewness	χ2 (*p*)	12.43 (0.26)	*27.09 (0.00)*	*22.1 (0.02)*	15.44 (0.12)	14.01 (0.17)
	Kurtosis	z (*p*)	-1.53 (0.13)	1.05 (0.29)	-0.06 (0.96)	-1.34 (0.18)	*-2.92 (0.00)*
Shapiro-Wilk		W (*p*)	0.99 (0.08)	*0.96 (0.02)*	*0.95 (0.01)*	*0.98 (0.03)*	0.992 (0.08)
Henze-Zirkler		Hz (*p*)	*1.18 (0.01)*	0.93 (0.07)	*1.03 (0.02)*	*1.12 (0.02)*	0.98 (0.11)
Royston		H (*p*)	3.88 (0.19)	4.67 (0.17)	*14.51 (0.00)*	5.67 (0.12)	5.45 (0.14)
PCA	Eigenvalue 1		21.57	11.77	11.03	15.04	21.20
	Eigenvalue 2		14.44	7.91	8.05	14.05	19.94
	Eigenvalue 3		7.43	3.00	4.86	5.43	9.13

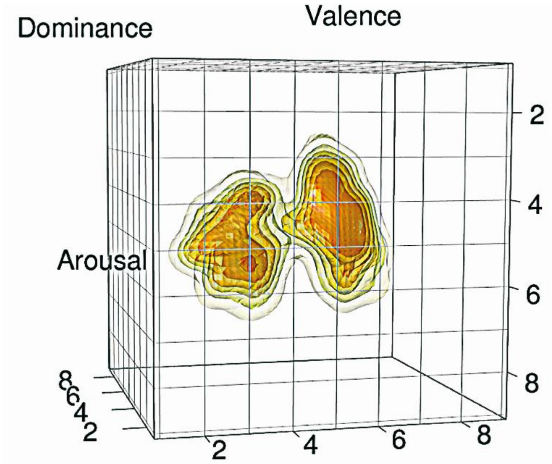
Note: Significant results indicating non-normality are printed using italic characters. Wherever rounding has resulted in near-zero values, these should be interpreted as $p<0.001$.

Table S3. BIC values for all the models considered within the model-based clustering procedure.

k	EII	VII	EEI	VEI	EVI	VVI	EEE	EVE	VEE	VVE	EEV	VEV	EVV	VVV
1	-8671.93	-8671.93	-8429.92	-8429.92	-8429.92	-8429.92	-6853.79	-6853.79	-6853.79	-6853.79	-6853.79	-6853.79	-6853.79	-6853.79
2	-7550.51	-7484.72	-7198.02	-7174.51	-7167.27	-7151.06	-6589.67	-6449.51	-6574.74	-6446.17	-6438.96	-6442.29	-6445.84	-6448.33
3	-7074.81	-7072.43	-6933.30	-6929.80	-6913.77	-6913.51	-6421.70	-6404.12	-6408.78	-6403.04	-6385.87	-6351.59	-6390.38	-6368.51
4	-6933.33	-6895.56	-6759.19	-6711.12	-6745.23	-6713.03	-6418.85	-6410.76	-6360.55	-6386.42	-6394.85	-6343.72	-6404.55	-6383.36
5	-6791.55	-6805.93	-6644.96	-6660.17	-6663.38	-6686.11	-6382.17	-6376.99	-6347.73	-6380.00	-6374.04	**-6341.11**	-6388.91	-6378.29
6	-6747.01	-6760.70	-6636.97	-6653.35	-6669.52	-6663.25	-6399.60	-6400.29	-6373.22	-6395.85	-6401.69	-6378.84	-6443.89	-6408.86
7	-6733.89	-6745.28	-6636.13	-6661.72	-6679.37	-6662.53	-6407.49	-6428.39	-6377.73	-6439.33	-6442.38	-6417.70	-6491.33	-6498.67
8	-6726.10	-6695.73	-6594.54	-6611.93	-6638.20	-6665.66	-6433.14	-6449.76	-6408.65	-6451.17	-6460.30	-6446.90	-6494.17	-6514.12
9	-6645.08	-6664.46	-6594.39	-6601.02	-6637.80	-6650.03	-6421.71	-6451.09	-6411.01	-6471.65	-6461.81	-6491.22	-6554.31	-6559.11

Note: According to Fraley et al. (2012), the abbreviations refer to the configuration of the clusters in a given model: EII = Spherical distribution, equal volume, equal shape, orientation not applicable; VII = Spherical distribution, variable volume, equal shape, orientation not applicable; EEI = Diagonal distribution, equal volume, equal shape, orientation on coordinate axes; VEI = Diagonal distribution, variable volume, equal shape, orientation on coordinate axes; EVI = Diagonal distribution, equal volume, variable shape, orientation on coordinate axes; VVI = Diagonal distribution, variable volume, variable shape, orientation on coordinate axes; EEE = Ellipsoidal distribution, equal volume, equal shape, equal orientation; EEV = Ellipsoidal distribution, equal volume, equal shape, variable orientation; VEV = Ellipsoidal distribution, variable volume, equal shape, variable orientation; VVV = Ellipsoidal distribution, variable volume, variable shape, variable orientation. Here, the BIC value underlined and in bold signals the optimal model, with *k*=5, and VEV as the most suitable configuration.

Figures



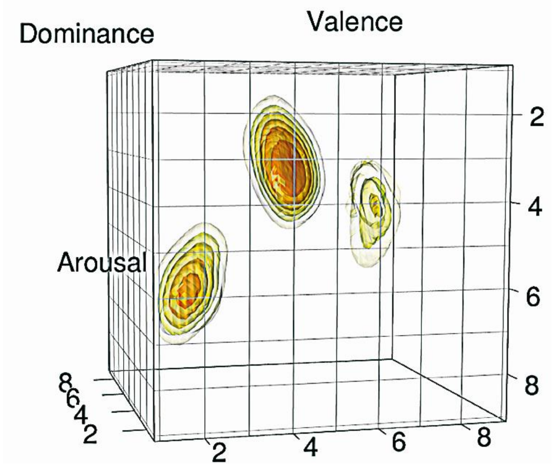


Figure S1. Density according to cluster, largely confirming the VEV model issued by Mclust(): varying volume/equal shape (ellipsoidal)/varying orientation. In the top section: clusters 1 and 5, in the lower section: clusters 2, 3 and 4.